

# Philips Technical Review

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

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

## EDITORIAL NOTE

Beginning with this volume, issues of Philips Technical Review will no longer be indicated by the month. The average size of the issues and the number of issues per year remain unchanged. The actual date of publication will appear on the first page of each issue.

## AN EMISSION ELECTRON MICROSCOPE FOR RESEARCH AT HIGH TEMPERATURES

by G. BAAS and G. W. RATHENAU \*). 537.533.35:621.385.833:620.186

*In the short time of its existence the electron microscope has already rendered inestimable service in many branches of science. Its importance as an aid to research can be judged from the numerous types of electron microscopes that have been developed in recent years. The emission microscope represents the outcome of a rather special line of development pursued at the Philips Research Laboratories in an attempt to apply electron microscopy to certain metallurgical investigations for which normal "transmission microscopes" are unsuitable.*

The emission microscope occupies a special place in the history of electron microscopy, for the first electron microscopes were of this type. These early microscopes were electron lenses which produced an image of an electron-emitting metallic surface. Strangely enough, however, the emission microscope has been put to practical use by only a few workers. About 1935 Burgers and Ploos van Amstel investigated nickel-iron, iron, zirconium, etc. with such a microscope, in the Philips' Research Laboratories at Eindhoven <sup>1)</sup>. In spite of the low magnification and resolving power of their microscope, they were able to derive valuable information on transformations in metals at high temperatures. Later, Mecklenburg and Kinder <sup>2)</sup> developed emission microscopes for

higher magnifications, but as far as we are aware, these were not employed in any systematic research and the development of microscopes of this kind was for a long time at a standstill.

There are perhaps two main reasons why the emission microscope has been neglected by designers of electron-optical instruments since the early days of electron microscopy:

- 1) From the theoretical investigations of Recknagel, et al., <sup>3)</sup> <sup>4)</sup> it became clear that the resolving power of any practical emission microscope must inevitably be far inferior to that of the "transmission microscope".
- 2) The transmission microscope has many applications in medicine, biology and other fields, that are virtually closed to the emission microscope.

\*) Professor at the University of Amsterdam; formerly of the Philips Research Laboratories in Eindhoven.

<sup>1)</sup> W. G. Burgers and J. J. A. Ploos van Amstel, Philips tech. Rev. 1, 312-316, 317-320, 1936; Physica 4, 5-14, 15-22, 1937 and 5, 304-312, 313-319, 1938; Nature 141, 330-331, 1938.

<sup>2)</sup> E. Kinder, Naturwiss. 30, 591-592, 1942.  
W. Mecklenburg, Z. Physik 120, 21-30, 1942/43.

<sup>3)</sup> A. Recknagel, Z. Physik 117, 689-708, 1940/41.

<sup>4)</sup> A. Recknagel, Z. Physik 120, 331-362, 1942/43. A useful review of the work of Recknagel and others on the emission microscope may be found in the article by H. Mahl; Die elektronenmikroskopische Untersuchung von Oberflächen, Ergebn. exakt. Naturw. 21, 262-312, 1945.

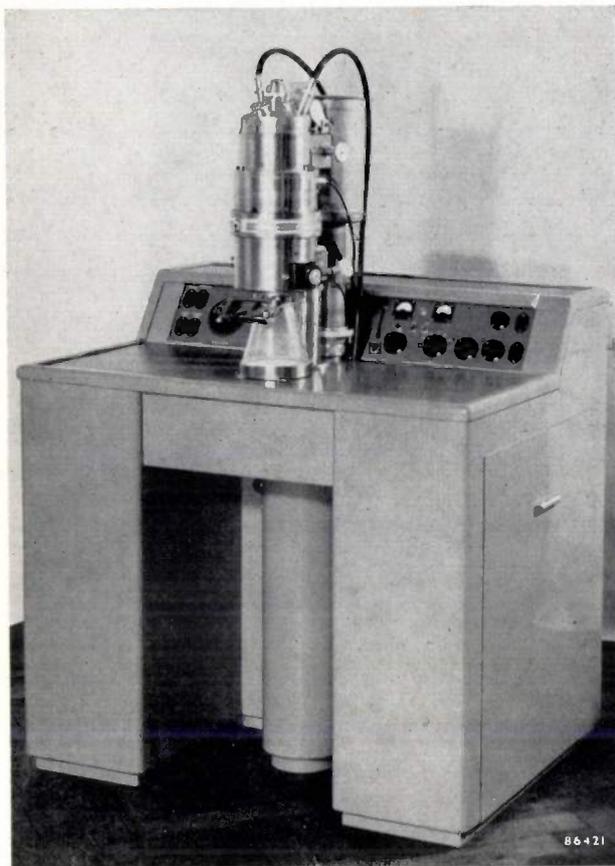


Fig. 1. Emission electron microscope developed by Philips.

We shall examine the first point in more detail later. As to the second point, it is also a fact that there are some fields of investigation where the emission microscope is pre-eminently suited<sup>5</sup>). This is especially true of the investigation of materials at high temperatures, already referred to. In polycrystalline metals, consisting of grains of like or unlike chemical composition, structures can be observed with the emission microscope that are unstable at room temperature. Moreover the grain growth and the transition of a metal into another phase can be observed while in progress, presenting a fascinating spectacle, which once witnessed is never to be forgotten.

After the war, at a time when great advances were being made in metal physics, and high-temperature materials were gaining in practical significance as a result of the development of new engines, plans were drawn up at Philips for an emission microscope for normal laboratory use. Direct impetus was given to this project by research into materials resistant to high temperature in connection with the Philips hot-air engine. The microscope developed in the

course of this work will be described in its present form (see *fig. 1*) in this article. The aim of the design was a higher magnification and resolving power than are possible with the optical microscope. The design is based on an emission microscope constructed to the specifications of Le Poole in the Philips Research Laboratories in Eindhoven, in 1947, for investigating oxide cathodes. Experience in the development of transmission microscopes was freely drawn upon, and consequently the microscope now to be described shows considerable similarity to the EM 75 KV transmission microscope, the subject of a recent article in this Review<sup>6</sup>).

#### Image formation in the emission microscope

In transmission electron microscopes, the object investigated is "illuminated" with a beam of electrons accelerated to a high energy. The electrons which, after passing through the object, carry the information and must be focused to an image, maintain an almost constant energy during the whole process. This is not the case in the emission microscope. Here the information is carried by electrons that have been emitted with very low velocity by the object under investigation, which is itself the cathode, and the focussing to an image must be accompanied by acceleration of the electrons. Hence the first (and sometimes the only) electronic lens in the emission microscope is always an electrostatic "immersion lens": the electrons are accelerated because the object itself lies in the electric field of the electronic lens. The "refractive index" in the space in which the electrons are propagated is therefore greater than 1 at the object. This is also a feature of immersion objectives in the optical microscope.

A direct consequence of the low velocity of emission of the electrons, is that electrons leaving the cathode at wide angles to the optical axis, are nevertheless sufficiently refracted by the field to be able to take part in the production of the image. This is another difference between emission and transmission microscopes, for in the latter, the paths of the electrons leaving the object make only very small angles with the optical axis.

These points of difference explain why, even on theoretical grounds, it is impossible to attain as high a resolving power with the emission microscope as with the transmission microscope. Practical limitations, as we shall see, make the situation still more unfavourable to the emission microscope.

<sup>5</sup>) See also: M. Gauzit and A. Septier, *Bull. micros. Appl.* 1, 109, 1951. R. D. Heidenreich, *J. appl. Phys.* 26, 757-766, 879-889, 1955.

<sup>6</sup>) A. C. van Dorsten and J. B. Le Poole, *The EM 75 kV, an electron microscope of simplified construction*, Philips tech. Rev. 17, 47-59, 1955/56.

An exhaustive study of the geometrical optics of the emission microscope by Recknagel<sup>3)</sup> has shown that the resolving power is given, to a first approximation, by

$$r = \frac{\epsilon}{E} \dots \dots \dots (1)$$

Here  $E$  is the strength of the electric field at the cathode surface, and  $\epsilon$  the most probable energy (in electron volts) of the electrons emitted by the cathode, which are assumed to have a Maxwellian velocity distribution.

Equation 1 indicates that for a high resolving power, the field strength  $E$  should be made as high as possible, and the emission energy  $\epsilon$  as low as possible. This is already understandable from physical considerations. Image defects (spherical and chromatic aberration) which limit the resolving power, will be smallest, the closer the electrons keep to the axis — and this requires both a high field strength  $E$  and a low initial energy  $\epsilon$ .

In practice the field strength  $E$  at the cathode is limited by the danger of electrical breakdown, which will be discussed more fully later. In the microscope to be described below, a field strength of 100 kV/cm (or, more exactly, between about 50 and 150 kV/cm) can be used. The mean emission energy  $\epsilon$  depends upon the temperature of the cathode. At a temperature of 1000 °C, such as obtains in the investigations under discussion,  $\epsilon \approx 0.1$  eV, so that according to equation (1) the maximum resolving power of this microscope will be 100 Å.

Recknagel has also studied the movement of electrons in the emission microscope from the standpoint of wave-mechanics<sup>4)</sup> and showed to what extent calculation of the resolving power by formula (1), derived from geometrical optics, is justified. Formula (1) is found to be valid, provided a magnitude  $\kappa_0$ , which represents, essentially, the ratio of  $\epsilon/E$  to the wavelength of the electrons at the cathode, is appreciably greater than 1 (*loc. cit.*, page 352).

According to the temperature of the specimen and the applied high voltage, the value of  $\kappa_0$  in the case in question lies between about 10 and 2.5, so that equation (1) is still applicable.

An important consideration in designing the present microscope, was the added and quite appreciable limitation of the resolving power, due to the imperfection of the observed surface under the conditions of investigation. Where the surface shows roughness — as most surfaces do when heated (temperature etching) — local deformations of the electric field occur, which act as extra lenses and

may impair the sharpness of the image at the points affected (*fig. 2*). Owing to this effect it is impossible to expect to be able to distinguish image details smaller than the order of 1000 Å. For most metallurgical investigations this is quite satisfactory: the resolving power is still several times higher than can be attained with the optical microscope. The above mentioned practical limitation, however, implies that there is no objection to supplementing the immersion lens by other lenses — even if the latter are attended by more pronounced image defects — should this be found useful or convenient for constructional reasons. Such extra lenses are used in the Philips emission microscope. Two magnetic lenses are in fact used, with which a total magnification of as high as 3000 × can be obtained without difficulty. The first magnetic lens operates in an unfavourable state of excitation, as will appear later, and is ultimately the factor that limits the resolving power of the microscope. The present resolving power is about 1000 Å.

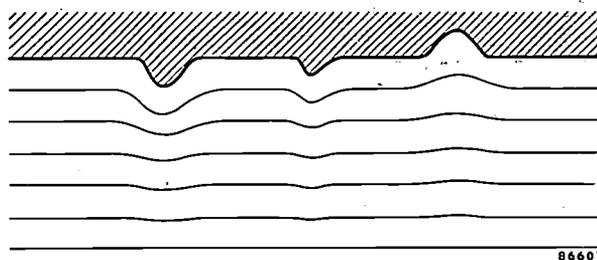


Fig. 2. Equipotential surfaces of the electric field which accelerates the electrons emitted by the specimen and functions as an electrostatic lens. Small elevations or pits in the surface of the specimen give rise to local deformations of the field, limiting the resolving power of the emission microscope.

### Contrast in the image

The information that the metallurgist seeks in the electron microscope image of an object is usually concerned with the grain structure and changes which occur when the temperature is varied. For this structure to be visible, not only must sufficient electrons be emitted by the metallic surface, but the emission must vary sufficiently from grain to grain. Sufficiently large differences in emission are not normally found. A homogeneous alloy, provided it emits sufficient electrons, will give an image merely on being heated, in which all manner of lines may be seen that appear to be grain boundaries (*fig. 3*); but such lines usually represent boundaries that had previously occupied those positions and no longer exist at the instant of observation.

A well-known artifice for producing a "contrast" between the grains, which at the same time makes for good emission at temperatures of practical interest, is employed in the Philips emission

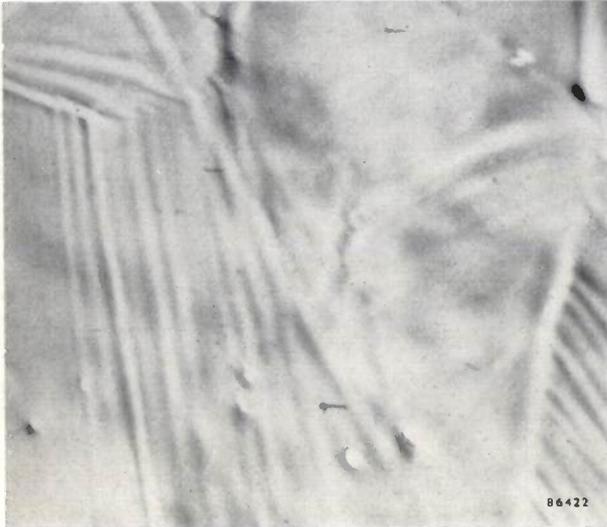


Fig. 3. Photograph of a homogeneous alloy with good thermionic emission. In this case the numerous lines, that might appear to be grain boundaries, are merely grooves in the surface at the locations of earlier grain boundaries.

microscope. It consists in the evaporation on to the surface under investigation of a very thin layer of a suitable metal (activator) which reduces the work function of the surface. The amount of activator adsorbed on to each crystallite surface depends on the orientation of the crystallite. Since the number of electrons emitted is determined almost exclusively by the number of activator atoms, differently oriented grains of a homogeneous material give different electron current densities. By the evaporation of activator atoms on to the surface, the latter is painted, as it were, in various shades of brightness. Moreover the high rate of migration of the adsorbed activator atoms over the surface makes it possible to follow changes in the crystalline structure continuously.

Thanks to the use of activators with very low work functions (alkali and alkaline earth metals) reasonable emission is obtained even at relatively low temperatures: with barium as activator the working temperature range is  $900^{\circ}$  to  $1200^{\circ}$  C, with caesium it is still lower, viz.  $500^{\circ}$  to  $740^{\circ}$  C.

The activators mentioned are exceedingly reactive, and because of this, care must be taken to exclude foreign elements from the vicinity of the object under observation. As a consequence, the vacuum is more critical in the emission microscope

than in the transmission microscope. Because of the reactivity of the activators this method of increasing contrast will not of course be applicable to all metals. Metals, having a high sulphur content, for example, do not lend themselves to this method of investigation, since the sulphur would react with the activator.

The cathode assembly of the emission microscope differs from that of the transmission microscope not only because of its different optical arrangement and the special provisions for the evaporation of activators, but also because the cathode must be adjustable through considerable distances during the observations, in order to be able to scan the object for interesting details. This is done by moving it in the plane at right-angles to the optical axis, just as in the transmission microscope. In the emission microscope, however a radically different construction is necessary because the object, i.e. the cathode, is at a high electrical potential.

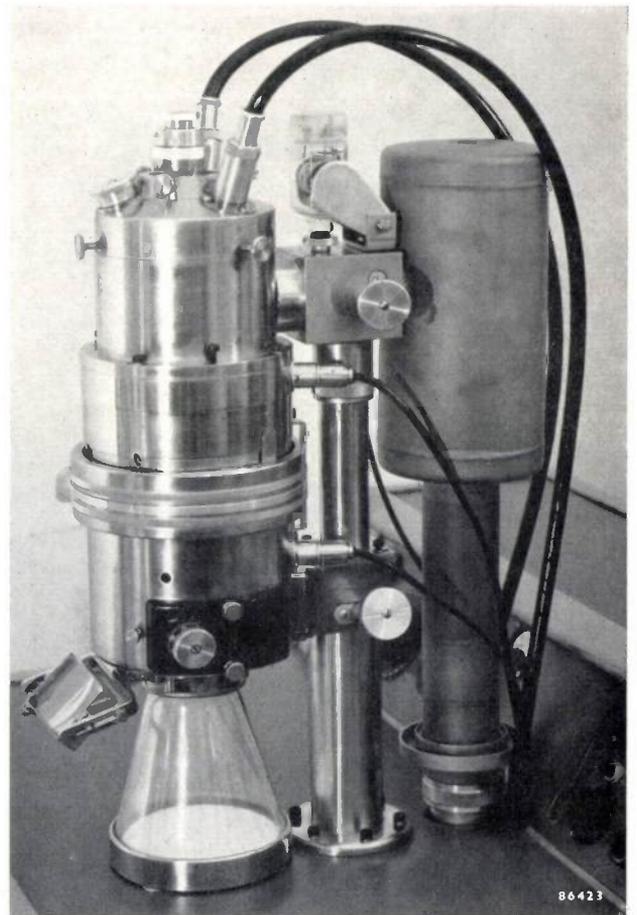


Fig. 4. Microscope tube of the emission microscope (left) and pumping tube with liquid air condenser (right). The lower part of the centre tube, which in the EM 75 kV is connected directly to the high-vacuum pump, now serves principally as a mechanical support fixing the microscope tube to the desk.

### Construction of the emission microscope

The vertical microscope tube, together with a part of the vacuum system, is shown in *fig. 4*. A cross-sectional diagram of the microscope tube is given in *fig. 5*. As already stated the design is modelled as far as possible on that of the EM 75 kV transmission microscope. Components peculiar to the emission microscope will be examined here more closely.

#### The cathode assembly

The cathode chamber in the upper end of the microscope tube is enclosed in an earthed steel jacket. The latter serves also to shield the specimen from any stray magnetic fields from outside, that would give rise to distortion. The low velocity of the electrons at the object, which is a feature of the emission microscope, makes such shielding essential: the deflection of electrons by a magnetic field at right-angles to the direction of movement is inversely proportional to their velocity.

The specimen itself, which is raised to a maximum potential of 45 kV with respect to earth, is mounted in a metal container suspended by three glass insulators from the cover of the cathode chamber (*fig. 6*). The lower part of the metal container forms a flat electrode, which can be regarded as a Wehnelt cylinder and has a small aperture immediately beneath the specimen, through which the electrons pass. The electrons then traverse the field between the Wehnelt cylinder and the earthed anode, which is fitted in the upper pole piece of the first magnetic lens.

For activating the surface of the specimen, four pairs of terminals are arranged about

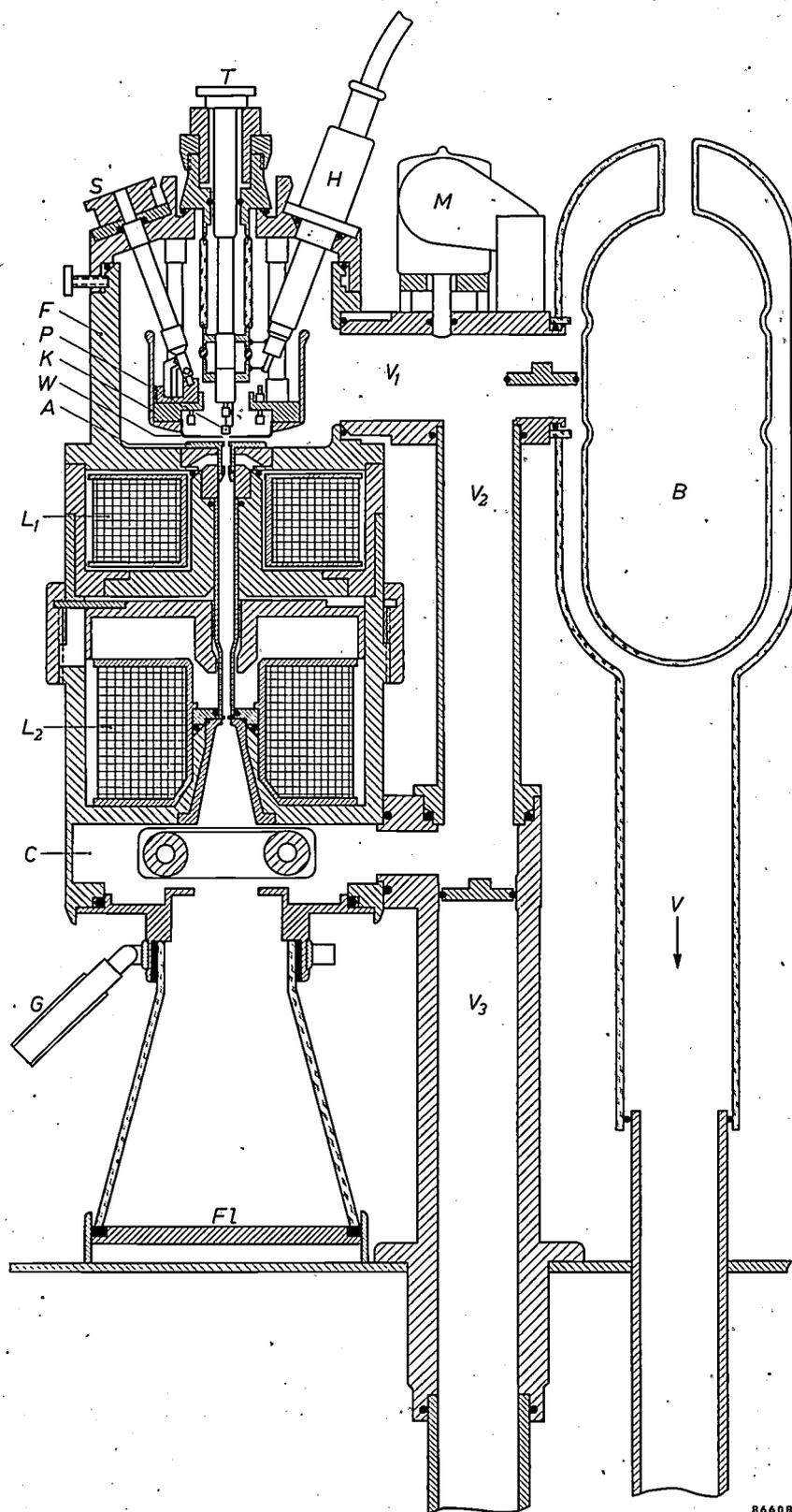
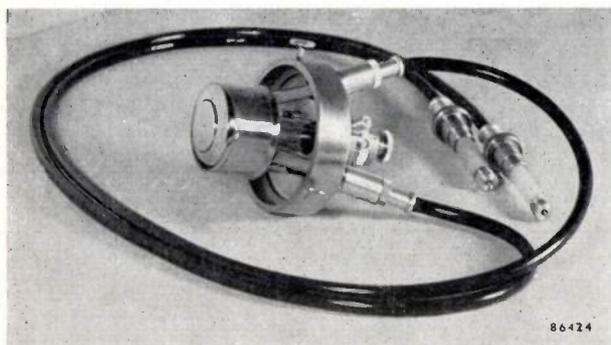
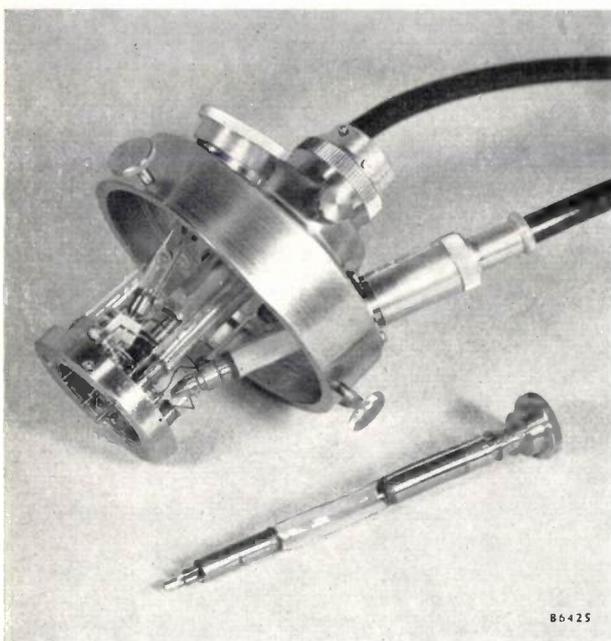


Fig. 5. Sectional diagram of the microscope tube (simplified). *K* cathode container, with the specimen *P* and the Wehnelt cylinder *W*, which are connected via the high voltage cable *H* to the negative H.T. terminal (maximum voltage 45 kV with respect to earth). *A* earthed anode; *F* steel jacket shielding the cathode chamber; *T* specimen holder; *S* switch for the activator heating filaments; *L*<sub>1</sub> and *L*<sub>2</sub> magnetic electron lenses; *C* camera chamber; *Fl* fluorescent screen; *G* glass observation lens; *V*<sub>1</sub> and *V*<sub>2</sub> pumping tubes communicating with the cathode chamber and camera chamber; *B* liquid air condenser; *V* tube leading to the high vacuum pump; *M* Philips manometer (Penning gauge).

the specimen in its metal container, to which coiled heating filaments can be attached. For activation with barium, a tantalum filament coated with barium beryllate is used. When the filament



86424



86425

Fig. 6. *Above*: Cathode chamber cover, with the insulated cathode container suspended from it and the high voltage cables (one for the specimen, one for the activators). The Wehnelt cylinder and the aperture through which the electrons make their exit may be seen.

*Below*: Cathode assembly with cathode container removed; the specimen holder has been taken out (foreground).

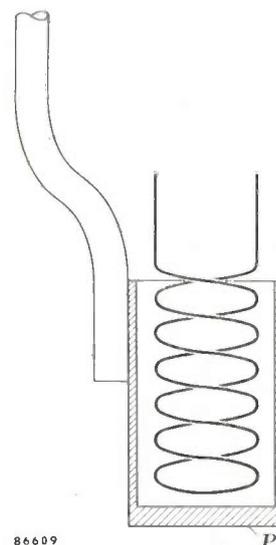
is heated the beryllate is reduced and the barium is given off as vapour. For activation with caesium, a mixture of caesium bichromate and silicon, packed into a small tube and heated by means of a filament, is employed. Other substances can of course be used as activators, e.g. strontium. The heating current for the filaments, which is drawn from a separate filament supply transformer with special high-voltage insulation, is fed to the cathode container via a separate high-tension cable and plug. An external switch makes it possible to select

the most suitable activator for the particular specimen and temperature range, while the microscope is actually in operation. Moreover as soon as one of the four activator sources is exhausted, it is possible to switch over to another.

When using barium, it is generally satisfactory to activate the specimen prior to observation. When caesium is used as activator it is necessary to keep the source switched on while observing, since the absorbed caesium quickly re-evaporates from the surface of the specimen.

The specimen to be investigated is fashioned in the shape of a small beaker, which can be indirectly heated by an alumina-coated coiled-coil filament placed inside it. This filament is wound non-inductively (as are the activation coils), so that the heater current does not give rise to an interfering magnetic field (see above). The surface to be observed, *P*, i.e. the base of the beaker (see fig. 7), is carefully polished to reduce the surface roughness effect described above to a minimum. A mounting bracket is then spot-welded onto the beaker in such a way that when fixed in the specimen holder, the surface *P* of the specimen lies exactly perpendicular to the optical axis. The specimen holder is clamped (but is free to move along its axis) in a glass cylinder, which can be moved laterally, i.e. at right-angles to its longitudinal axis by means of adjusting screws on the outside. The electrically insulated specimen holder, which can be adjusted from the outside via vacuum sealed glands can be rotated about its longitudinal axis and also translated along this axis during observations.

Since the Wehnelt cylinder and its exit aperture remain in the same position during all these manipulations, the lateral displacement and the rotation make it possible to scan a large area of the object surface. This method of scanning, which is possible only with the emission microscope, is in practice far more convenient than the conventional displacement of the object in two directions at right-angles to each other: far fewer movements are



86609

Fig. 7. The metallic surface to be investigated *P* forms the base of a small beaker, in which a heating filament is inserted. The beaker is attached to the specimen holder by means of a rod welded onto the former.

necessary for scanning the whole surface (only a small part of which is visible on the screen) and it is considerably easier to re-locate a particular part of the object. The displacement of the specimen holder along the axis permits of accurate regulation of the distance between the object and Wehnelt cylinder; this is of importance in the formation of the image; see below. When one specimen is to be replaced by another, the specimen holder is simply taken right out.

During axial adjustment of the position of the object, it is unnecessary to interrupt the supply of high voltage and heater current to the specimen, but when one specimen is replaced by another the supply must be temporarily interrupted. For this purpose two sleeves, insulated from each other, are mounted on the object end of the specimen holder; when the latter is inserted in the glass cylinder, the sleeves make contact with two contact springs. These contacts are connected via flexible wires to the end of the high-voltage cable in the cathode chamber (see the cross-sectional diagram of the cathode assembly in fig. 8).

The temperature of the specimen can be determined quite accurately by means of a thermo-couple or a pyrometer, using a dummy specimen mounted and heated in the same manner, but not raised to a high potential.

#### The magnetic electron lenses

In the emission microscope the role of objective lens and specimen stage in the transmission microscope is taken over by the magnetic lens  $L_1$  (fig. 5). To reduce image defects it would be preferable to energize the lens fairly strongly; see the article cited in <sup>6)</sup>. To do this, a rather novel lens construction with a very wide bore and a very large number of ampere-turns would be needed, in view of the special form of anode necessary to produce an electrostatic accelerating field of the required homogeneity <sup>4)</sup>. However to bring the lens construction of the emission microscope more in line with those used earlier in Philips electron microscopes, a lens of relatively narrow bore has been made, which has a sufficiently short focal length and can be sufficiently energized for normal focusing merely by virtue of a somewhat unusual shape of its upper pole piece. In spite of its considerable image aberrations, this lens is quite adequate for the emission microscope, for reasons given earlier.

A second magnetic lens ( $L_2$  in fig. 5), which is practically identical to the projector lens of the EM 75 kV, raises the magnification to the desired value, and also provides a simple means of adjusting

the magnification. This is done by varying the gap between the upper and lower pole pieces of the projector lens by means of a milled ring encircling the microscope tube as in the EM 75 kV <sup>5)</sup>. In this

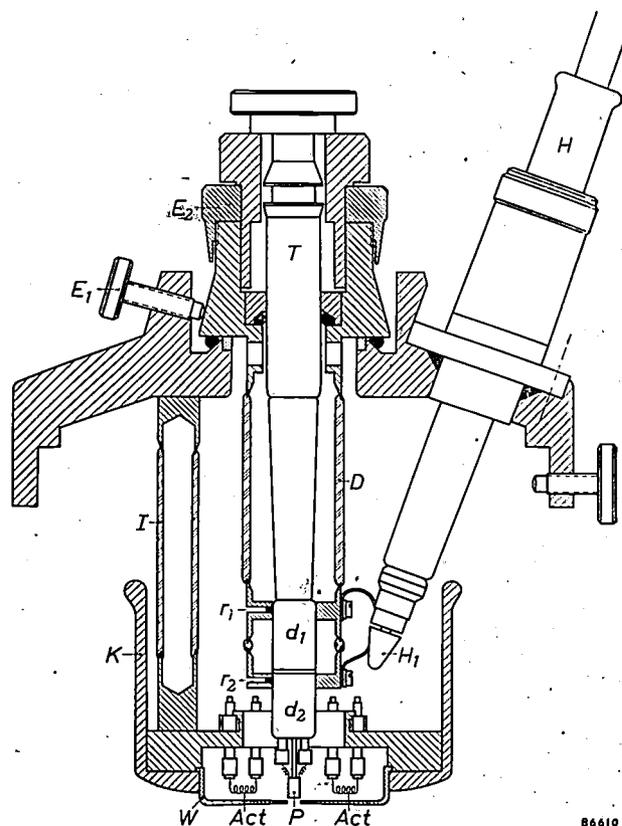


Fig. 8. Cathode assembly (simplified: the sections of the left and right halves do not lie in the same meridian plane).  $I$  glass insulators;  $D$  glass cylinder into which the specimen holder  $T$  carrying the specimen  $P$  is inserted. By means of screws  $E_1$  the glass cylinder can be displaced in a plane perpendicular to its longitudinal axis. With the screw  $E_2$  the specimen holder is raised or lowered in the direction of its longitudinal axis. The specimen is connected to the high voltage plug  $H_1$  via the insulated sleeves  $d_1$  and  $d_2$  and the sliding contacts  $r_1$  and  $r_2$ .  $Act$  activator filaments, with separate high tension lead for the heater current. Remaining letters as in fig. 5.

way the magnification of the lens can be varied continuously between 20 and 90 (i.e. a range of nearly 5; the device employed in the EM 75 kV for increasing this range up to a factor of 9, is unnecessary here).

The total magnification is usually 150 to 750, the first magnetic lens contributing a factor of 10 and the electrostatic immersion lens a factor of approximately 1. This is when the specimen is in its lowest position i.e. as near as possible to the "Wehnelt cylinder". If the distance is increased by raising the specimen (screw  $E_2$  in fig. 8), while at the same time the magnetic lens is more weakly energized, the total magnification increases, attaining a value of 3000 in the extreme case, when the first magnetic

lens is completely unenergized. This is illustrated in *fig. 9*, in which the image formation is depicted under various conditions.

A magnification of 3000 is sometimes useful for visual observation. For photographic work a magnification of 750 is more suitable, with the specimen at its lowest position and with relatively powerful excitation of the first magnetic lens, since under these conditions the field strength at the cathode is greatest (see equation 1). At this magnification the grain size of the photographic emulsion in no way detracts from the resolving power, so that the latter can be fully exploited by subsequent optical enlargement of the photographed image.

As with the EM 75 kV, photographic records are made with a 35 mm film holder mounted in the vacuum space beneath the second magnetic lens. With this arrangement the progress of a phenomenon can be recorded in about 40 exposures. If necessary a phenomenon can be filmed with a ciné-camera, mounted beneath the fluorescent screen at the base of the conical glass end of the microscope tube. The normal metal fluorescent screen must then be replaced by a glass screen, permitting observation of the image from below.

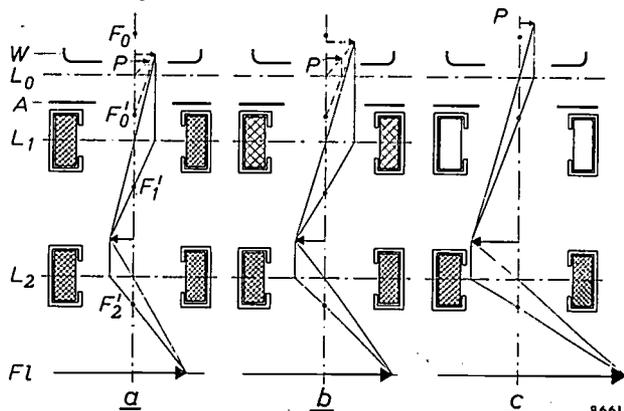


Fig. 9. Image formation in the emission microscope, for different magnifications. (Rotation of the image by the magnetic lenses has been disregarded.)

a) The specimen  $P$  is as close as possible to the Wehnelt cylinder  $W$ . The electrostatic lens produces a somewhat enlarged virtual image, which is situated behind the cathode.

b) As the specimen is drawn back somewhat, the image falls back and becomes bigger. The focal length of the first magnetic lens must now be increased, i.e. the lens more weakly energized.

c) When the distance between the specimen and the Wehnelt cylinders attains a given value, the virtual image lies at infinity and changes abruptly into a real image at infinity on the other side, i.e. in front of the cathode. This image approaches the cathode rapidly as the object is moved back still further and soon lies in the object plane of the second magnetic lens. The first magnetic lens is then completely unenergized.

Owing to the very large depth of focus of the electron microscope and the fact that the concentration of the rays is in itself incomplete (limited resolving power), the transition from (b) to (c), surprisingly enough, is fairly continuous in practice.

### The vacuum system

It will perhaps already be clear that in spite of certain similarities, the emission microscope is not so simple to operate as the transmission microscope EM 75 kV. This follows from the very nature of the investigations for which the emission microscope is employed: in these investigations the specimen itself plays an essential role in the optics of the microscope. Many of the simplifications in the EM 75 kV have therefore not been used in the emission microscope in cases where a more complex construction was desirable in the interests of efficiency.

This is especially evident in the vacuum system. On heating a specimen there is always the possibility of rapid out-gassing with the risk of electrical breakdown; measures must therefore be taken, firstly to get rid of any liberated gases as quickly as possible and secondly to minimize the effects of a possible breakdown. These problems are resolved by inserting a liquid-air condenser ( $B$  in *fig. 5*) in the vacuum tube just where the latter enters the cathode chamber, which serves to remove any condensable gases. Moreover a mercury diffusion pump is used for production of the required high vacuum. In the EM 75 kV an oil diffusion pump is employed, because, to mention one reason, production of satisfactory high vacua with a mercury pump would entail incorporating a liquid air condenser on account of the vapour pressure of mercury. In the emission microscope, where the condenser is in any case indispensable, the mercury pump is to be preferred, because it precludes the formation of oil films in the cathode chamber, and the deposition of graphite from residual oil vapour should electrical breakdown occur. A rotary oil pump is used as the backing pump. There is no necessity for this pump to run continuously while the microscope is in use, since a vacuum reservoir tank is incorporated; this tank is conveniently accommodated in the desk designed for the EM 75 kV. The mercury pump requires a backing pressure of a few mm of Hg. A mercury switch brings the rotary oil pump in operation automatically, as soon as the backing pressure rises above the permissible value.

It has been found desirable to separate the cathode chamber and the camera chamber as completely as possible; for even after outgassing at  $50^\circ\text{C}$  in vacuo for several hours, photographic film still gives off minute amounts of substances that combine with the activator on the surface of the specimen rendering it ineffective. Within the microscope tube, the cathode chamber and the camera chamber are fairly well separated by the

narrow apertures of the two magnetic lenses. Outside the microscope tube, there is ready access between these two chambers via the two tubes that open into them ( $V_1$  and  $V_2$  in fig. 5); but the junction of  $V_1$  and  $V_2$  is located as near as possible to the liquid-air condenser, so that the condensable gases from the film are unlikely to gain access to the cathode chamber. While the microscope is in use a pressure of between  $10^{-5}$  and  $10^{-6}$  mm of Hg is measured at this junction with a sensitive Philips manometer (Penning gauge). (The main function of the vacuum tube  $V_3$  in fig. 5, which in the EM 75 kV leads to the high-vacuum pump, is, in this instrument, to provide the microscope tube with a single robust attachment to the desk. However,  $V_3$  is used at the same time for preliminary evacuation of the microscope tube by the backing pump.)

To avoid electrical breakdowns or, at worst, to minimize the effects of such a breakdown, all surfaces in the cathode chamber are chromium plated, while various components, such as the anode and the Wehnelt cylinder, are made from stainless steel and carefully polished. Such surfaces can easily be cleaned and give off little gas. The anode and the aperture of the first magnetic lens, which is most openly exposed to contamination with vaporized activator, are in one piece; this component can be easily screwed out of the upper pole piece for cleaning.

#### Stabilization of the high voltage

Stabilization of the accelerating voltage, which always requires care when magnetic electron lenses are employed, is more difficult in the emission microscope than in most transmission microscopes: the investigation of specimens at differing temperatures means that the electron emission and therefore the flow of electrons through the microscope tube must be able to vary within wide limits, without substantially affecting the potential on the tube.

A stabilizer suitable for this instrument has been designed by S. van der Meer of the Research Laboratories in Eindhoven, and is illustrated in the simplified diagram in fig. 10. The high voltage generator  $B$ , supplied by the mains, delivers a maximum direct voltage of about 45 kV. Fluctuations in the output voltage  $V_B$  of  $B$  which occur as a result of rapid mains voltage fluctuations, etc., are almost fully compensated by the output voltage  $V_A$  from a D.C. amplifier  $A$  connected in series with  $V_B$ . The principle of this control circuit has been discussed several times in this Review<sup>7)</sup> and is briefly recapitulated in the caption to fig. 10.

This control circuit in itself is not enough because the voltage may vary considerably: changes in load and fluctuations in the mains voltage can give rise to a variation in  $V_B$  of several kV (3 kV on passing from no load to full load), if no precautions are taken. The amplifier  $A$ , of course, is unable to deliver such a high compensating voltage  $V_A$ , and an additional control is therefore incorporated

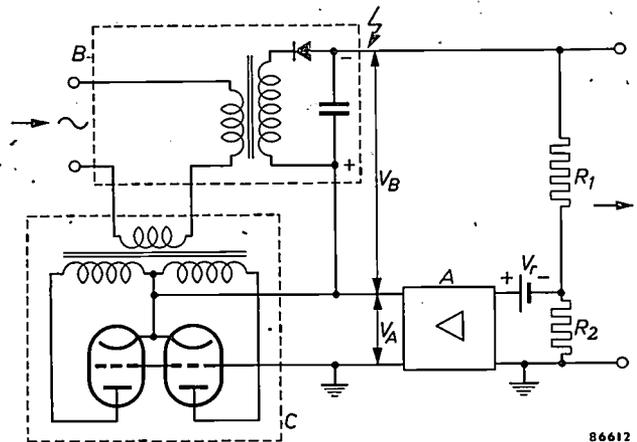


Fig. 10. Circuit diagram of H.T. voltage regulator.  $B$  high tension generator, with output voltage  $V_B$ . A fraction of this voltage, nominally equal to the reference voltage  $V_r$  (a few tens of volts) delivered by a battery, is tapped off through the voltage divider  $R_1$ - $R_2$ , and the difference between  $V_r$  and the voltage tapped off is supplied to the D.C. amplifier  $A$ . The output voltage  $V_A$  fluctuations in which are opposite in sign to the input voltage, is in series with  $V_B$ , so that rapid variations in  $V_B$  are compensated by opposing variations in  $V_A$ . The unavoidable residual variations of  $V_B + V_A$  are very small, provided the amplifier  $A$  has a sufficiently high amplification factor. The variable impedance  $C$ , controlled by  $V_A$  and regulating the input voltage to the high voltage generator, compensates large slowly-occurring variations in  $V_B$ , resulting from load variations and mains voltage fluctuations.

in the circuit in the form of a variable impedance  $C$ , connected in series with the primary side of the high voltage generator  $B$ . The impedance is so controlled by the voltage  $V_A$ , that variations in the input voltage to the high voltage generator compensate for the variations in  $V_B$ , and the amplifier  $A$  can never be overloaded. Control via  $C$  has a somewhat delayed action, but this is no objection, since large variations, such as occur on raising the cathode temperature of the microscope, are always fairly gradual. The rapid fluctuations, on the other hand, are never large and can therefore be met by control  $A$ .

The relative constancy of the high voltage obtained with this circuit is better than 1:3000 within half a minute. This is adequate for the purpose.

<sup>7)</sup> See A. C. van Dorsten, Philips tech. Rev. 10, 135-140, 1948/49.

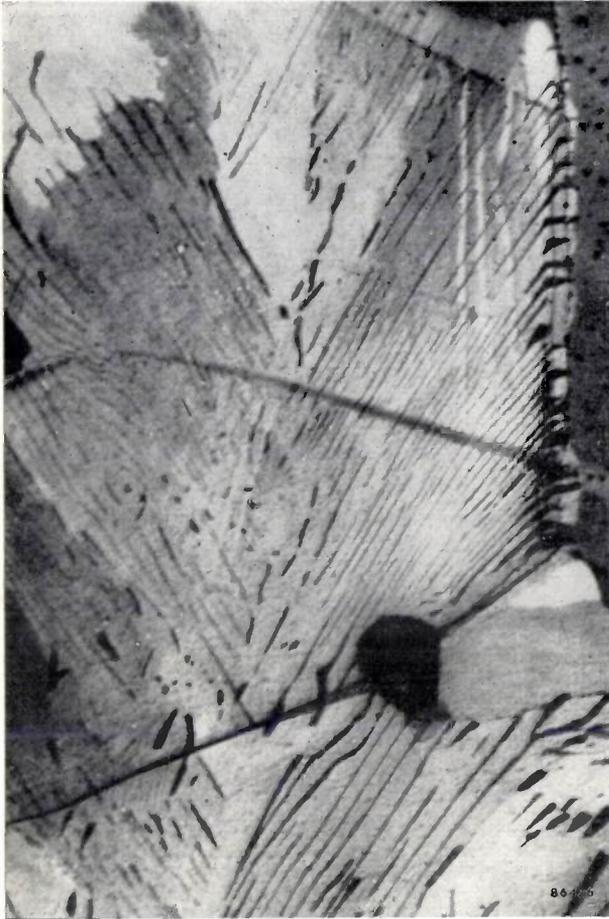


Fig. 11. Photograph taken with the aid of the Philips emission microscope showing the transition of carbon steel from austenite into pearlite. Magnification about  $2500\times$ . The resolving power here is somewhat better than  $1000\text{ \AA}$ .

The energizing currents for the magnetic lenses must also be stabilized. This is achieved by means of a circuit described earlier <sup>8)</sup>.

<sup>8)</sup> Philips tech. Rev. 12, 33-51, 1950/51.

A number of reports on investigations carried out with the emission microscope described above, have been published, one having appeared in this journal <sup>9)</sup>. A photograph made with the present microscope is shown in fig. 11 and gives an idea of the image quality obtained.

The assistance of A. C. van Dorsten and his co-workers A. Verhoeff, A. J. J. Franken and J. E. Jansz, all of the Eindhoven Research Laboratories has been of great value in the development of this microscope.

<sup>9)</sup> G. W. Rathenau and G. Baas, *Physica* 17, 117-128, 1951; *Métaux* 29, 139-150, 1954; *Acta Met.* 2, 875-883, 1954; G. Baas, Demonstration of the austenite-pearlite transformation by means of the emission electron microscope, *Philips tech. Rev.* 16, 337-339, 1954/55.

**Summary.** Description of an emission electron microscope, having a resolving power of about  $1000\text{ \AA}$  and an (adjustable) magnification of up to  $3000\times$ . After a brief review of the theoretical and practical limitation of the resolving power by electron-optical considerations, the nature of the objects to be examined is discussed in greater detail. Since this microscope is primarily intended for the investigation of metals at high temperatures, thermionic emission is used. Activators such as barium or caesium evaporated on to the surface of the object, ensure adequate emission in temperature ranges of practical interest, and likewise produce differential emission from the variously oriented grains of a polycrystalline metal (contrast). The construction of the cathode permits the use of four different activators while the microscope is actually in use, scanning of wide areas of the object surface and variation of the degree of magnification by adjusting the field of the first electron lens (an electrostatic "immersion lens") Apart from this lens, the electron-optical system employs two magnetic lenses, the second of which is practically identical with the projector lens of the transmission electron microscope EM 75 kV and, as in the latter, permits variation of the magnification. The emission microscope resembles the EM 75 kV, in many other details of its construction. Nevertheless there are fundamental differences in the vacuum system and the system for stabilizing the accelerating voltage of 45 kV. The demands made on both the vacuum system and the voltage stabilization are far higher for the emission microscope than for the transmission microscope.

## LIGHTING IN TRAINS AND OTHER TRANSPORT VEHICLES WITH FLUORESCENT LAMPS

by L. P. M. ten DAM and D. KOLKMAN.

621.327.534.15:628.977.8:629.1.06

---

*The tubular fluorescent lamp is a light-source of high efficiency and long useful life. These are indeed the main reasons for its large-scale adoption for the lighting of houses, shops, offices, streets, and so on. One particular field into which the fluorescent lamp has quite recently been introduced is in the lighting of land transport vehicles and small ships. The storage battery which serves as the source of energy in these cases will of course deliver direct current and will also have a rather low voltage. Lamps and auxiliary equipment therefore had to be specially developed for this field of application.*

---

The power for electric illumination in trains, buses and small ships is supplied by storage batteries, charged by a dynamo. Since the batteries can deliver only a limited amount of energy, lighting in these vehicles was rather poor while designers were restricted to the use of the filament lamp.

An improvement became theoretically possible when some 20 years ago a new light-source with a considerably higher efficiency came on the market — the fluorescent gas discharge lamp. It was, however, not possible to use it at that time in trains and other transport vehicles, since the first lamps of this type were developed exclusively for alternating current. Direct current lamps did appear later, but their ignition voltage was higher than the voltage of the battery equipment normally used in trains, ships, etc. Only recently have fluorescent lamps and ballast equipment been developed for battery supplies of comparatively low voltages. Such lamps are the Philips "TL" C lamps, operating on 72, 100 and 100 V.

Another system which is used where the voltage of the battery is lower than 72 V (e.g. 24 V) is the conversion of the direct current into A.C. with a voltage of about 220 V. In this case normal lamps can be used, but special ballast equipment is necessary for reasons which will appear below.

**"TL" C lamps with current regulator, operating on D.C.**

Storage batteries such as are used in vehicles show considerable variation in voltage; thus the lighting batteries in trains whose nominal voltage is, for example, 72 or 110 V, may vary from 60 to 105 V or from 80 to 150 V respectively. Fluorescent lamps fed directly from these batteries must be so constructed that they will ignite faultlessly at the lowest voltage of the battery, i.e. 60 V or 80 V. We shall see later how this is achieved.

Apart from the ignition voltage, the working voltage is one of the characteristics of the lamp which can be specified within close limits during design. From the point of view of efficiency it is advisable to choose the working voltage as close as possible to the battery voltage: the loss in the stabilizing resistor — the ballast which must always be included with D.C. supply — is then kept to a minimum. The working voltage of a gas-discharge tube such as the "TL" C type is virtually independent of the current and by using a ballast resistor whose resistance is also independent of the current it is a simple matter to calculate the current through the lamp at all values of battery voltage. This current always varies relatively more than the battery voltage, and the closer the working voltage of the lamp is to the battery voltage, the greater the variations in current. For mains operation, where voltage fluctuations have a maximum of  $\pm 10\%$ , a good stability can be obtained by making the working voltage approximately equal to half the nominal mains voltage. The loss in the resistor is then about the same as the power consumption of the lamp itself. This means that with a 72V battery the lamp should have a working voltage of 36 V. Calculation shows that a lamp which consumes 15 W at the nominal battery voltage uses only 10 W at 60 V, but shows the high figure of 29 W at a voltage of 105. The light output of the lamp is approximately proportional to the consumption, and will thus vary by a factor of nearly 3. If the working voltage is placed higher in order to limit the loss in the resistor, e.g. equal to  $\frac{2}{3}$  of the nominal battery voltage (i.e. 48 V), then the light variations will become even greater; they can in such a case attain a factor of 4.

In practice wide variations in light output are found to be very disagreeable. When selecting the working voltage, a compromise must therefore be

sought between the need to achieve high efficiency and that of limiting light variations. This compromise can be favourable affected if in place of a resistance of constant value a current regulator tube (barretter) is used as ballast; in this way the value of the resistance is automatically varied so that the current remains more or less constant during fluctuations in voltage. Very constant current can be achieved with a resistor consisting of an iron filament in a hydrogen-filled bulb<sup>1)</sup>. Such a tube, filled with an inflammable gas, would not be permissible in a public transport vehicle however, and moreover the iron filament is not shockproof.

Both these objections have been overcome in the current regulator tubes which have been developed for our purpose. The pure hydrogen is replaced by a mixture of hydrogen and nitrogen, and the iron filament by a coiled tungsten filament; the limits of the current variation are, it is true, not so narrow as those obtainable with iron-hydrogen resistors, but the gas is not inflammable and the filament is stronger. The stabilizing effect is improved and the shock resistance increased yet further by winding the coil on an insulating tubular support (fig. 1). Fig. 2 shows the characteristic of one of these current regulator tubes which have been developed for train lighting; the current is given as a function of the voltage across the tube.

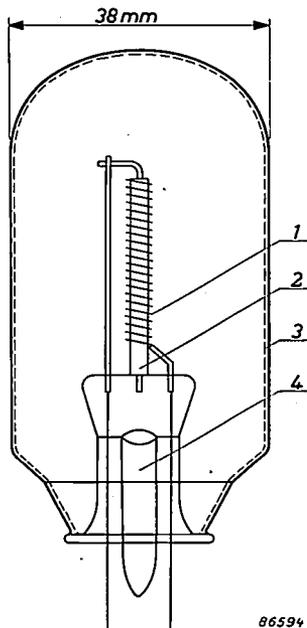


Fig. 1. Cross-section of a current regulator tube for "TL" C lamps. 1 tungsten filament, wound on an insulating pillar 2. 3 glass bulb. 4 pumping stem. Gas filling: non-inflammable mixture of hydrogen and nitrogen.

1) J. G. W. Mulder, Philips tech. Rev. 3, 74-79, 1938.

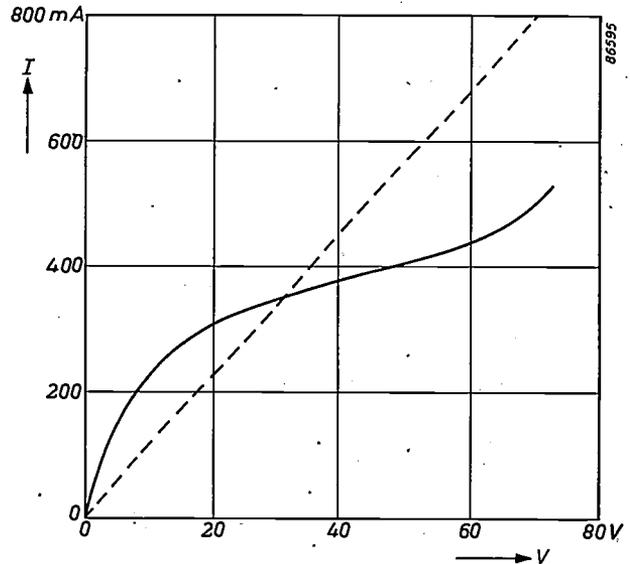


Fig. 2. Solid line: characteristic of one of the current regulator tubes for "TL" C lamps: current  $I$  as a function of the voltage  $V$  across the regulator tube. For comparison, the dotted line shows the characteristic of a constant resistance.

The use of current regulator tubes has made it possible to set the working voltage of the fluorescent lamps somewhat higher than half the nominal battery voltages, while keeping the variations in light output below a factor 2. Table I summarizes the main details of these "TL" C lamps for 14, 15 and 20 W.

The efficiency at a battery voltage of 72 V is about 25 lumens per watt. With incandescent lamps it is only about 10 lm/W; in this connection it should be noted that incandescent lamps require some type of electro-mechanical regulator to prevent overloading, and that a certain loss must occur in this. Apart from the higher efficiency, the "TL" C lamp has the advantage that the electro-mechanical regulator can be replaced by a tube.

Table I: Data on "TL" C lamps for lighting in trains.

Power consumption (nominal) W	Battery voltage (nominal) V	Current (nom.) mA	Working voltage V	Length inches	Diameter inches
14	72	370	38	15	1½
15	72	335	45	18½	1½
20	110	355	56	24½	1½

Ignition of the "TL" C lamp

To guarantee ignition at the lowest voltage delivered by the batteries the "TL" C lamp is fitted with special electrodes and connected in a circuit as shown fig. 3. With the lamp extinguished, the contact of push-button  $D_1$  is open and that of push-button  $D_2$  closed. When light is required  $D_1$  is

depressed and the current flows via  $D_1$  through the starting resistor  $R_1$  (contained in the bulb of the current regulator tube) and the lamp cathode  $K$ . The latter consists of a coiled tungsten filament which

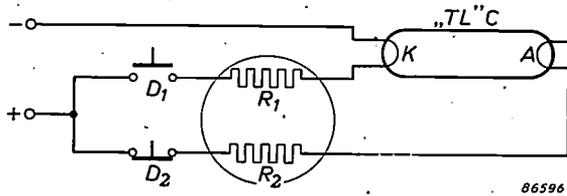


Fig. 3. Circuit of a "TL" C lamp, with cathode  $K$  and anode  $A$ .  $D_1$  "on" switch.  $D_2$  "off" switch.  $R_1$  starter resistor.  $R_2$  ballast resistor (current regulator tube).  $R_1$  and  $R_2$  are mounted in the same bulb.

is longer than that for an A.C. lamp rated for the same current. Thus a sufficiently large part of the battery voltage appears across the cathode for an arc to develop between its extremities. To facilitate the formation of this arc small metal plates are fixed on either side of the cathode (fig. 4), which

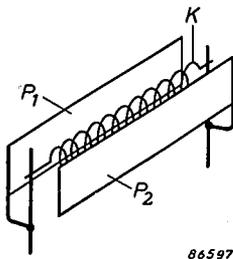


Fig. 4. Small metal plates  $P_1, P_2$  mounted on either side of the cathode  $K$  of a "TL" C lamp. These plates assist in the formation of the starting arc across the cathode, and also prevent the blackening of the ends of the tube.

also serve to prevent blackening of the envelope at the ends of the lamp. The gas in the arc across the cathode is ionized so that the main discharge via the current regulator tube  $R_2$  and anode  $A$  is

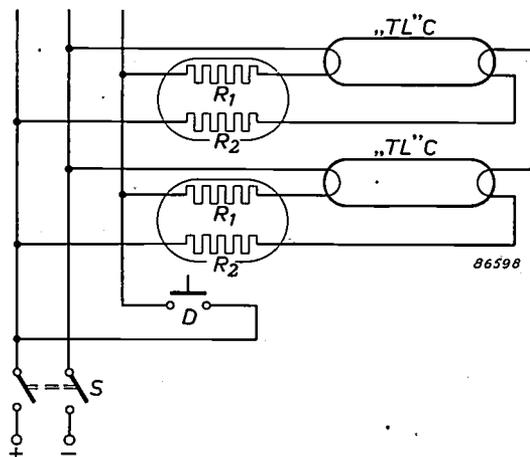


Fig. 5. Circuit for the simultaneous switching of a number of "TL" C lamps, all provided with a starter resistor  $R_1$  and a ballast resistor  $R_2$ . After closing switch  $S$  and depressing button  $D$  all the lamps will light ( $D$  may be a time-delay switch which automatically cuts out the cathode heater current after a certain period). Switch  $S$  is used to extinguish the lights.

easily ignited. (To promote ignition of the relatively long "TL" C 20 W lamp — see Table I — this type has two conductive strips attached to the inside of the glass, each being connected to one electrode.) As soon as the main discharge has started button  $D_1$  is released and the lamp remains alight. To extinguish the lamp it is only necessary to depress button  $D_2$ .

Electrode  $A$  of the lamp is exactly similar to  $K$ , making the lamp completely symmetrical.

If it is unnecessary to be able to switch each lamp on and off separately, the simple circuit of fig. 5 can be used (see caption). Combinations of both circuits are also possible so that certain lamps (in the passenger compartments) can be switched on and off separately, and others (in the corridor and toilets) only as a group.

An example of a train installation using "TL" C lamps is illustrated in fig. 6.

#### A.C. lamps supplied via a D.C./A.C. converter

On many railways the battery voltage is 24 V, and in buses the starter battery, also usually 24 V, is used for the lighting. It is not possible to manufacture fluorescent lamps of high efficiency for this very low voltage. In these cases recourse is made to the second solution mentioned above: the D.C. is converted into alternating current and the voltage is then stepped up through a transformer to a value of about 220 V, when standard A.C. lamps can be used.

#### The centrifugal converter

There are several methods of converting direct current into alternating current. The most obvious one (a motor driving a generator, i.e. a rotary converter) has the disadvantage of a low efficiency, which largely offsets one of the advantages of fluorescent lighting. The vibrator-converter (generally used for car radios) seems attractive by reason of its high efficiency but is not suitable for the power output which is needed here — viz. some few kVA. Similar in principle, but able to cope with the required output, is the centrifugal converter (fig. 7)<sup>2)</sup>. In this, a rotating jet of mercury makes alternate contact with stationary segments and thus performs

<sup>2)</sup> H. Böhm, Turbowechselrichter mit Quecksilberstrahl für die Stromversorgung von Zugbeleuchtungsanlagen mit Leuchtstofflampen, AEG-Mitt. 42, 255-259, 1952.  
H. Böhm, Stromrichter mit umlaufendem Quecksilberstrahl, E.T.Z. A 74, 478-480, 1953.  
Another type of D.C.-A.C. converter suitable for this purpose is described in: K. Schmidt, Spezial-Umformer für Leucht- röhren in Fahrzeugen, Elektro-Technik (Würzburg) 37, 79-81, 1955 (No. 10).

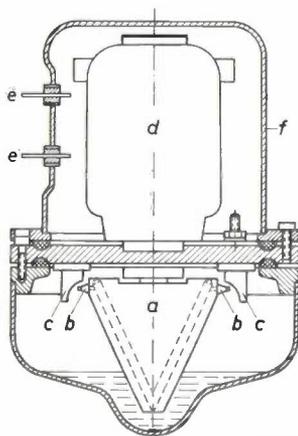


Fig. 6. A restaurant car of the Compagnie Internationale des Wagons-Lits, lit with 15 W "TL" C lamps.

an analogous function to the vibrating contact spring of a vibrator converter. The switching occurs in a sealed steel casing containing a reducing gas at a pressure of more than 1 atmosphere. A small electric motor driving the jet is contained in a second casing also filled with reducing gas but free of mercury. These measures make very little maintenance necessary.

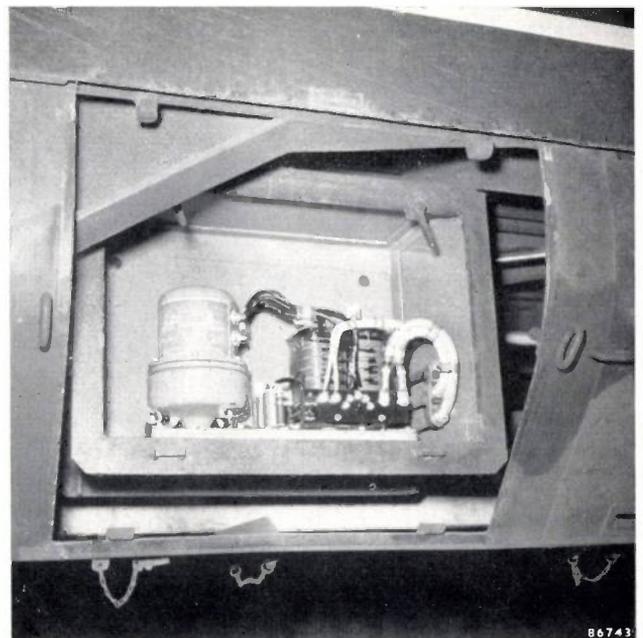
The centrifugal converter is available for outputs from 0.3 to 3 kVA; the photograph of fig. 7 shows the model for 1.5 kVA. The maximum efficiency with the

smallest type is about 85% and with the largest type about 90%. The frequency of the delivered alternating voltage depends upon the speed of rotation of the driving motor. A nominal frequency in the neighbourhood of 100 c/s (but not higher) should preferably be chosen, or, alternatively, 64 c/s. The frequency of 64 c/s has the advantage that if connected to a normal mains of 220 V, 50 c/s, the lamp current will be the same as with the converted voltage of 250 V, 64 c/s delivered by the converter; thus, during cleaning or repair of the vehicle in the



86599

Fig. 7. Centrifugal converter (made by AEG, Germany). *Left*: Cross-section (taken from the article in the E.T.Z. referred to in <sup>2</sup>). *a* rotating cone. *b* nozzles for the two mercury jets (two to give dynamical balance). *c* stationary electrodes. *d* driving motor, with gas-tight current connections *e* through the casing *f*. *Right*: Photograph of a 1.5 kVA centrifugal converter unit.



86743

workshop or depot, the lighting can be connected directly to the mains and not run off the battery (the battery would soon be drained if it supplied lighting with the engine not running). Advantages of 100 c/s are that flicker is less than at 64 c/s and the ballast can be made smaller and lighter. Working at frequencies considerably higher than 100 c/s leads to constructional difficulties in the centrifugal converter.

#### Ballast equipment for installations with centrifugal converter

The A.C. side of the centrifugal converter behaves in several respects differently from an A.C. mains network:

- The output voltage is not sinusoidal but of square waveform.
- The amplitude and frequency of the alternating voltage vary with the battery voltage (and in the same sense as the latter).
- The switching effect of the rotating jet of mercury implies that voltage and current must pass through zero simultaneously. If the load is such that, with a sinusoidal A.C. supply (of the same voltage curve as that of the converter) the current and voltage do not pass through zero together (power factor  $< 1$ ), then supply from the converter will cause a transient to occur at the beginning and end of each half cycle. This will cause losses and thus reduce the efficiency.

The ballast equipment must meet the following requirements:

- 1) The lamps should ignite at the lowest battery voltage (80% of the nominal value) and give sufficient light, while not being overloaded at the highest battery voltage.
- 2) The lamps should not ignite before the electrodes are heated up (otherwise the electrodes will be unduly damaged each time current is switched on).
- 3) It should be possible to switch off a large number of the lamps without the remainder being overloaded.
- 4) The power factor should be approximately 1 when supplied from mains with the same voltage curve as the converter.

We shall consider separately the case where all the lamps are switched on or off together, and the case where it is required that some of the lamps may be controlled individually. Consider first the case where there is only one lamp.

*One lamp.* Where "TL" lamps are supplied from A.C. mains the ballast usually consists of a choke in

series with the lamp. Where supply is from a centrifugal converter the use of this otherwise attractive circuit is not feasible because of the low value of the power factor (see above, sub-section c). It is not possible to improve matters with a capacitor across the input terminals, on account of the presence of strong harmonics in the alternating voltage delivered by the converter.

A circuit giving good results is that of Steinmetz (fig. 8) which was discussed not long ago in this Review<sup>3</sup>). When supplied with a sinusoidal alternating voltage this circuit has the property that the

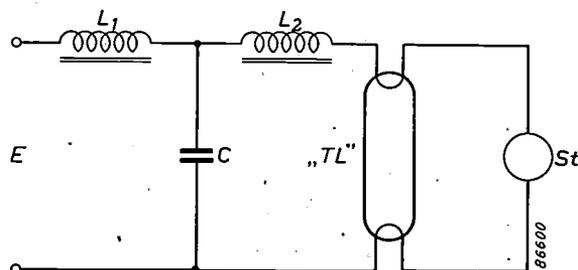


Fig. 8. "TL" lamp (with starter St) in a Steinmetz circuit. If  $L_1$  and  $C$  are in resonance with the frequency of the sinusoidal alternating voltage  $E$ , and  $L_2 = L_1$ , the power factor is unity.

power factor is 1 if the self-inductance  $L_1$  and the capacitance  $C$  are in resonance with the mains frequency and  $L_2$  is equal to  $L_1$ . As will be seen below, however, it is desirable in our case to deviate from the condition of resonance.

Some measurements were made with the circuit shown in fig. 8, supplied from a centrifugal converter whose battery voltage  $E_b$  and speed of rotation could be varied independently of each other. The rotation speed determines the fundamental frequency  $f$  of the alternating voltage. At various speeds  $E_b$  was adjusted so that the lamp current remained constant. Fig. 9 shows the measured values of the alternating voltage  $E$  and the power factor  $\cos \varphi$  as a function of  $f/f_0$ , where  $f_0$  is the resonant frequency  $1/(2\pi\sqrt{L_1 C})$ .

Bearing in mind the necessity for a constant lamp current, it is best to place the working point on the rising portion (*a-b*) of the curve for  $E$ , i.e. well removed from the resonance ( $f/f_0 = 1$ ); in this region, for constant lamp current, decreasing alternating voltage is indeed accompanied by decreasing frequency, just as in practice a decrease in battery voltage will produce a reduction in both frequency and alternating voltage. It would thus be possible to make the speed of revolution of the motor dependent upon the battery voltage in such a way that the lamp current will remain practically con-

<sup>3</sup>) W. Elenbaas and T. Hehenkamp, Philips tech. Rev. 17, 288-293, 1955/56 (No. 10).

stant over a wide range when this voltage drops. We see, however, in fig. 9 that the power factor in the range  $a-b$  is relatively low. For this reason,

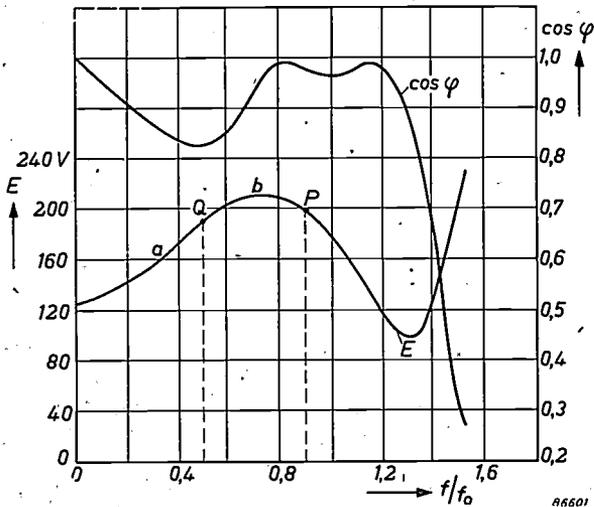


Fig. 9. Measured values of the alternating voltage  $E$  (r.m.s. value) and the power factor  $\cos \varphi$  in the circuit shown in fig. 8 supplied by a centrifugal converter as functions of the speed of revolution, i.e. of the frequency  $f$ . ( $f_0$  = resonant frequency of  $L_1-C$ ). The battery voltage was so adjusted that the current through the lamp remained constant (420 mA).

the working point has been placed to the right of the maximum but still off resonance, approximately at  $P$  ( $f/f_0 = 0.9$ ). It is true that the lamp current is then not constant, but it remains within admissible limits. This has been found from measurements, the results of which will be given presently.

*Many lamps, all switched on and off together.* When a number of lamps  $n$  are controlled by a single switch, it is sufficient to make use of a common choke with self-inductance  $L_1/n$  and a common capacitor with a capacitance  $nC$  where  $L_1$  and  $C$  are the values required for one lamp; each lamp has only one individual choke  $L_2$  (fig. 10). This arrangement gives a considerable economy in components.

There is one drawback which would be serious in the case of single lamps (fig. 8): although  $L_1$  and

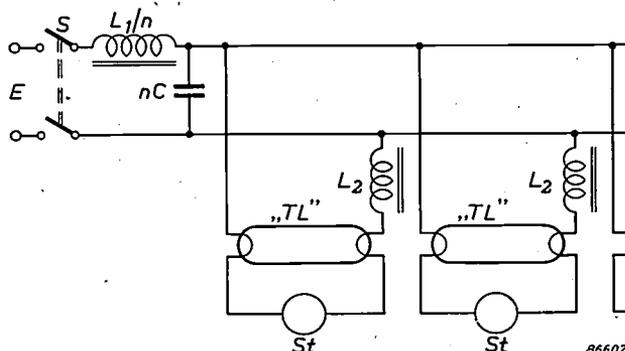


Fig. 10. Circuit for simultaneous control of a number of "TL" lamps  $n$ . The common choke has a self-inductance of  $L_1/n$ , the common capacitor a capacitance of  $nC$ . Each lamp has its own choke  $L_2$  and starter  $St$ .

$C$  are not in resonance with the frequency of the converter, the voltage across  $C$  after switching on, which is also across the still unlit lamp, is so high that the lamp ignites immediately, i.e. before the starter has had time to operate. This ignition with cold electrodes results in a shortening of the life of the lamp. In an installation with a large number of lamps, however, (fig. 10), one or other of the lamps will ignite with cold electrodes when switch  $S$  is closed, but as soon as one or two lamps are alight the capacitor voltage has dropped so much that the remainder of the lamps ignite only after a certain delay, via the starter, i.e. with hot electrodes. Some damage is caused, but only to the one or two lamps which ignite first. The disadvantage is therefore relatively less important than in the case of one lamp.

*Many lamps, some switched individually* If it is necessary for some of the lamps to be switched on and off separately (e.g. in railway compartments), it is undesirable, as we have seen above, to connect them as in fig. 8, since the lamps would then suffer too much when switched on. It is better to arrange them as in fig. 11 where each lamp has its own choke

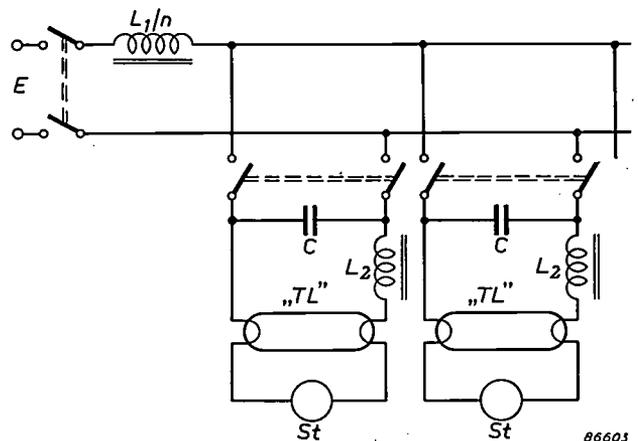


Fig. 11. Circuit for the individual switching of "TL" lamps. Common choke with self-inductance  $L_1/n$  (for a total of  $n$  lamps); each lamp has its own capacitor  $C$ , choke  $L_2$  and starter  $St$ .

$L_2$  and its own capacitor  $C$  (switched in and out with the lamp), and the  $n$  lamps have a common choke with a self inductance  $L_1/n$ . When all lamps are switched on the total capacitance is thus  $nC$ ; the values  $L_1$  and  $C$  are again so chosen that  $f/f_0 = 0.9$  (fig. 9) and the working point thus lies at  $P$ . If now the lamps are switched off one by one, then each time the remaining capacitance is decreased by an amount  $C$ , and  $f_0$  becomes higher as a result; the working point shifts along the  $E$  curve of fig. 9 to the left. This continues satisfactorily until about 70% of the load is switched off. If still more lamps

are switched off, a strong fifth harmonic occurs in the lamp current, overloading the remaining lamps. It is therefore necessary for at least 30% of the load to remain in circuit; this is no drawback in a railway carriage since these lamps can be used for lighting the corridor and toilets. When only this basic lighting is burning, the capacitance in circuit is 0.3 nC, giving a resonant frequency of  $f_0' = f_0/\sqrt{0.3}$  so that  $f/f_0' \approx 0.5$  (point Q in fig. 9). The fact that the power factor is rather low with this value of  $f/f_0'$  is now of less importance, since the converter carries a very light load.

For the switching on of the basic load, a similar situation exists as when all the lamps are switched together (see above): only the two lamps which ignite first suffer any damage. Once the basic load is in circuit, individual lamps which are switched on subsequently ignite via the starter, i.e. with pre-heated electrodes.

**Results**

Fig. 12 shows various electrical and luminous quantities as a function of the battery voltage. These were measured on an installation with a centrifugal converter and "TL" lamps of 40 W,

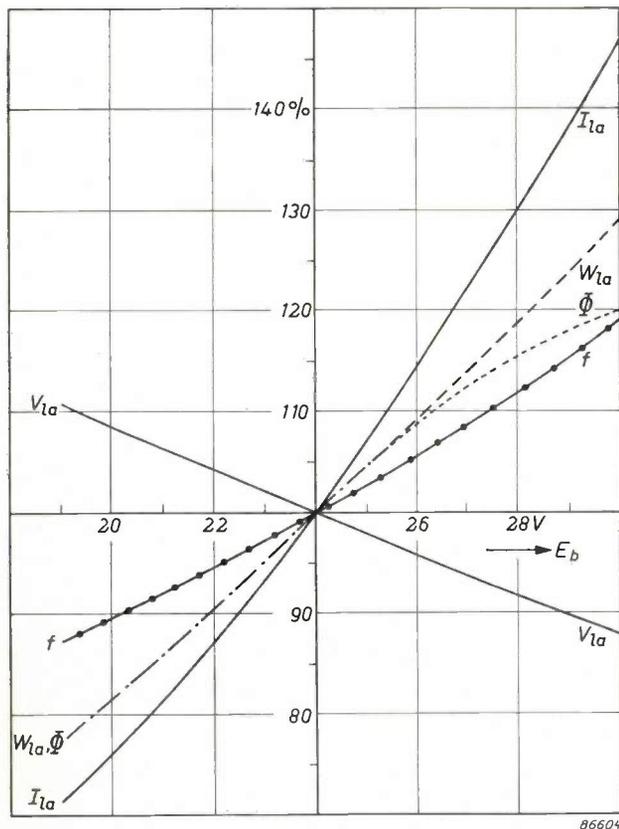


Fig. 12. Various quantities as a function of the battery voltage  $E_b$  in an installation with 40 W lamps supplied via a centrifugal converter: lamp current  $I_{la}$ , working voltage  $V_{la}$ , lamp dissipation  $W_{la}$ , luminous flux  $\Phi$  and the frequency  $f$ .



Fig. 13. Lighting in the carriage of an express train of the Netherlands Railways, using 40 W "TL" lamps, fed via a centrifugal converter.

colour "warm white" (colour temperature 3000 °K). The nominal battery voltage is 24 V, the minimum 19 V. The power consumption of the lamp at 19 V is 77.5% of that at 24 V; to make comparison possible it may be noted that the same lamp in a normal circuit with ballast choke and fed from mains with a sinusoidal alternating voltage and constant frequency would consume only 64% of the nominal power with the same reduction in voltage.

With reference to the output from the battery, the efficiency reaches a value of 45 lumens per watt. With incandescent lamps, as already mentioned, the efficiency is not more than 10 lumens per watt.

With an ambient temperature of 20 °C and a battery voltage of 80% of the nominal value, all lamps ignite reliably. At 0° C ignition can be guaranteed down to 83% of the nominal battery voltage. Fig. 13 shows a lighting installation using the centrifugal converter.

**Starterless circuit**

Not long ago a new fluorescent lamp, type "TL" M, with a starterless circuit, was discussed in this Review<sup>3</sup>). This lamp can also be used in

conjunction with the centrifugal converter, in the circuit shown in *fig. 14*. The difference between this and the normal "TL" M lamp circuit is that in the

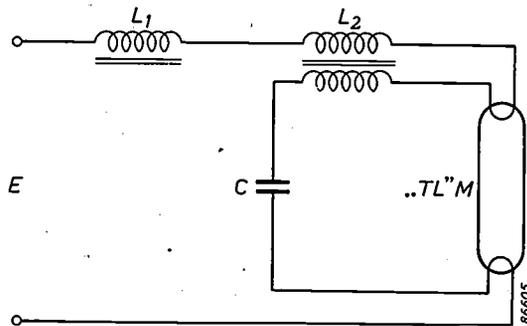


Fig. 14. Circuit for a "TL" M lamp, without starter. On switching on, there is an initial current through  $L_1$ ,  $L_2$  (whose coils are so wound that the total self-inductance of  $L_2$  is zero), the two filaments and  $C$ . Since  $L_1$  and  $C$  are close to resonance, the current is considerable and the filaments are thus rapidly heated. When they have reached a certain temperature, the lamp ignites. The currents through the two windings of  $L_2$  are now out of phase, and a certain self-inductance occurs; the circuit then effectively reverts to that of *fig. 8* (without the starter).

latter the choke coil  $L_1$  is omitted and its function taken over by magnetic leakage from the coils  $L_2$ . In the present case however it is necessary to retain choke  $L_1$  to achieve a good power factor. By making this common to all lamps in the installation,

lifetimes remain satisfactory, even with individual switching.

As a result of the extra losses in the electrodes and the coils the efficiency of an installation of "TL" M lamps is some 15% lower than that of an installation with starters; against this there is the advantage of simplification attending the use of "TL" M lamps, which need no starters.

**Summary.** During recent years more and more use has been made of tubular fluorescent lamps for the lighting of trains, buses and small ships. Special types ("TL" C lamps) have been developed for battery supplies of 72, 100 or 110 V using a current regulator tube (tungsten filament in a non-inflammable mixture of hydrogen and nitrogen) as ballast. The working voltage of the "TL" C lamps is rather more than half the nominal battery voltage; the regulator tube limits current fluctuations due to varying battery voltage to a factor of 2. "TL" C lamps are manufactured for power ratings of 14, 15 and 20 W. With a battery voltage of 72 V the efficiency is in the region of 25 lumens per watt.

When using 24 V batteries (e.g. the starter battery of buses) the direct current is converted into A.C. by means of a centrifugal converter (rotating mercury jet) of high efficiency, and the A.C. voltage raised to 220 V through a transformer, making it possible to use normal "TL" lamps. The frequency should preferably be either 64 or about 100 c/s; flicker, especially with the latter frequency, is practically nil. The ballast equipment (differing according to whether the lamps are switched on individually or together) is based on a Steinmetz circuit. The efficiency, with reference to the output from the battery, is as high as 45 lumens per watt. "TL" M lamps (starterless lamps) can also be used for illumination in transport vehicles.

## A PRIVATE AUTOMATIC BRANCH EXCHANGE USING HIGH-SPEED UNISELECTORS

by B. H. GEELS \*).

621.395.264

---

*In large automatic telephone exchanges economy demands that the number and dimensions of the necessary equipment be as small as possible without limiting the facilities available to subscribers. Only with the development of high-speed uniselectors and their associated test-relays, replacing the conventional two-motion selectors, has it been possible to utilize fully the advantages of the so-called indirect systems. The description of a large private automatic branch exchange (PABX) given in this article will elucidate this matter and will also give an idea of the elaborate measures necessary in order to meet all kinds of secondary requirements in a private telephone system. This article will be followed in due course by a description of the high-speed uniselectors and of a public exchange developed along the same lines.*

---

The traffic handled by private branch exchanges can be divided into two main classes, viz. the internal traffic and the incoming or outgoing traffic via the public telephone exchanges. Although the requirements made upon private exchanges as regards internal connections are generally less stringent than those made upon public telephone exchanges (it is not necessary, for instance, to count the calls), the establishment of connections for incoming and outgoing calls via the public lines makes very high demands on the private branch exchange. The development of PABXs has therefore followed entirely different lines from that of public exchanges.

For manual private exchanges there was no reason for any special development: an efficient telephone operator could be expected to cope with the demands of the incoming and outgoing traffic. Although automatic equipment has now been widely adopted for private traffic, the well-known manual exchanges equipped with flexible cords, jacks and plugs are still frequently used for external connections via the public line.

Although manual exchanges with jacks for every extension and for every public line offer the advantage of individual handling of every external call, they nevertheless have many disadvantages. Their operation makes considerable demands on the operator, as she must act promptly both at the beginning and at the end of each call. It may be necessary to employ a large number of operators, if the traffic is very heavy. Finally, the arrangement of the desk becomes confused in large installations on account of their very size. For these reasons the modern development of private branch exchanges

aims at the introduction wherever possible of automatic devices for the establishment of local calls whereby the demands made on the operator are reduced. In this way the cost of operators is cut to a minimum, and in addition local calls are handled in a much shorter time.

The development of the PABX type UB 49 is also based on this principle. By the use of various modern components developed by Philips Telecommunications Industries since the war, it has been possible to incorporate in this exchange a number of important features<sup>1)</sup>.

In order to appreciate these features it is necessary to discuss the development of automatic telephone equipment and in particular that of private automatic branch exchanges. The function of elements such as final selectors, group selectors, line finders, pre-selectors, etc., which form the component units of all automatic telephone exchanges is briefly explained in the diagrams in fig 1a and 1b for those readers who are not acquainted with these matters.

### Direct systems

The first automatic telephone system was developed by the American Strowger. The systems derived from this original Strowger system, which have been very widely applied, are called *direct systems*, because the dial pulses from the caller's extension are used directly for the positioning of the automatic telephone selectors<sup>2)</sup>.

\*) Philips Telecommunication Industries, Hilversum, The Netherlands.

<sup>1)</sup> A description of the PABX type UB 49 by the same author appeared in *Communication News* 15, 2-19, 1954 (No. 1).

<sup>2)</sup> See for example, J. Atkinson, *Telephony*, Pitman, London 1952.

Direct systems have the advantage of a relatively simple design and for a long time it was considered that they were the fastest type of system. Direct positioning of selectors ensures that the connection is established as soon as the subscriber has dialled the last digit of the required number and practically immediately afterwards the ringing current is applied to the called person's telephone. On closer examination, however, it will be found that the rapid action of these systems is illusory, as the telephone dial functions relatively slowly. The selectors of the direct systems are only just capable of following the dialling pulses, but as the subscriber does not realize the fact that time also passes during the dialling operation, he gets the impression that the establishment of a connection takes no time at all. If the modern selecting methods were to be used, however, the selectors of the direct systems would no longer be able to follow the selecting signals, which are much faster in the modern systems.

The nature of the direct systems is such that a number of devices are necessary for positioning the selectors and, after having established the connec-

tion, remain engaged although inactive during the entire call. Consequently, a relatively large number of these devices are necessary in the exchange and it is difficult to design them so as to be cheap to make.

The selector normally used in direct systems is a modified version of the Strowger two-motion selector (*fig. 2*). It has 100 outlets, so that 100 subscriber's lines can be connected to it. The selector must then react to two series of dialling pulses i.e. two digits of the required number. In order to make all the 100 outlets accessible in this way they are arranged in 10 rows of 10. The brushes of the selector can be moved by means of two electro-magnets which may be termed the vertical magnet and the rotary magnet. The pulses relating to the first digit are passed to the vertical magnet via the contacts of relays forming part of the selector circuit. The magnet lifts the brushes step by step till they face the row of outlets corresponding to the first digit. Some relay contacts are then tripped, so that the dialling pulses of the second digit are fed to the rotary magnet. The brushes then rotate step by step in a horizontal plane touching the contacts of the row as they pass, until they connect with the contacts of the desired subscriber. Since the positioning mechanism is discontinuous, the not inconsiderable mass of the brush mechanism must be accelerated and stopped again and again. The pulse repetition frequency, therefore, must not be too high: otherwise the selector will not be able to follow the pulses. For this reason a pulse repetition frequency of only 10 pulses per second was used in most direct systems.

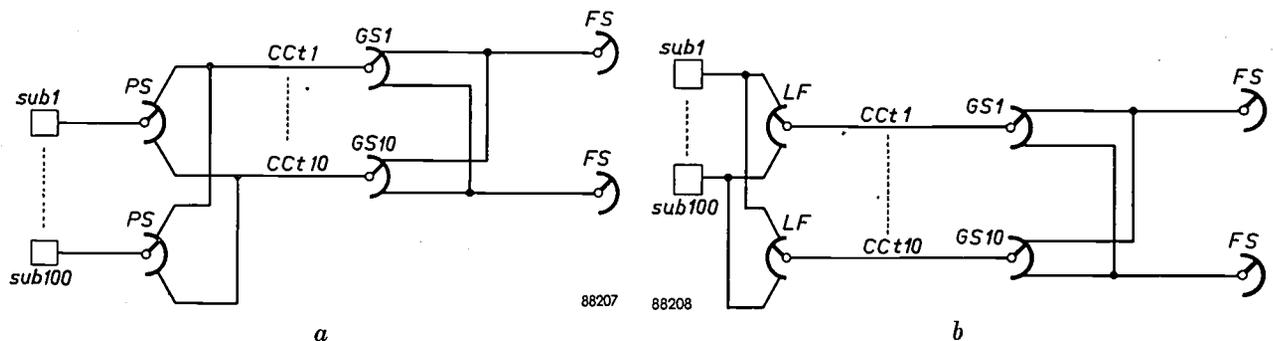


Fig. 1. Schematic diagrams illustrating automatic telephony. The extensions (sub) connected to a telephone exchange are mostly divided into groups of one hundred, which are then connected to the contact banks of selectors having 100 outlets (final selectors *FS*) to enable them to be selected.

If the exchange is equipped for 100 to 1000 subscribers the final selectors *FS* are preceded by the group selectors *GS*; in this way it is made possible to choose a free final selector to the desired group of 100 subscribers. In the case of 1000 to 10,000 subscribers it is required to introduce another group selector stage preceding the first one in order to make it possible to select a free group selector to the desired group of 1000 subscribers.

Consequently, to establish the desired connection the subscriber must have a group selector at his disposal. Only a small number of subscribers make a call at the same moment; it is, therefore, wasteful to connect every set to the brushes of a group selector. The subscriber's lines are therefore connected to a considerably smaller number of group selectors via one or more reducing stages consisting either of pre-selectors or of line finders. The number of group selectors depends on the average number and duration of the calls in the busy hour. It may be necessary, for example, to provide a group of 100 subscribers with 10 group selectors.

a) If pre-selectors (*PS*) are used as a reducing stage, then the contacts of each pre-selector are connected to the brushes of the 10 group selectors via cord circuits (*CCt*), while the

brushes of each pre-selector are connected to a subscribers' line. In this case 100 pre-selectors with 10 outlets are thus required for a group of 100 subscribers. If a subscriber lifts the receiver, then the brushes of the pre-selector relating to this subscriber start hunting the contacts; they stop at the contacts leading to the first free group selector. A relay, the test relay, which is consecutively connected to the various group selectors by the pre-selector, is provided for checking whether a group selector is engaged or disengaged. If the test relay is connected to a disengaged group selector, then it is energized, on account of which the pre-selector comes to a standstill. (In order to prevent existing calls via engaged group selectors from being disturbed by the testing, this is done not via one of the two speech lines, but via a third line and a third contact bank of the pre-selector.)

b) If line finders (*LF*) are used as a reducing stage, their contacts are connected to the subscriber's lines, while their brushes are connected to those of the group selectors. Consequently in this case 10 line finders with 100 contacts per bank are required. If the receiver is now lifted, the brushes of a free line finder start rotating along the contacts in order to hunt the calling subscriber. For this purpose a test relay connected to a brush of the line finder is connected in turn to the line circuit of the various subscribers. As soon as it is connected to that of the calling subscriber, the relay is energized and in this way the line finder is brought to a standstill.

**Indirect systems**

In *indirect systems*, as the name implies, the principle of direct control of selectors by the dialling pulses is no longer applied. The dialling pulses are instead received by a central switching device (register or director) which is connected to the cord

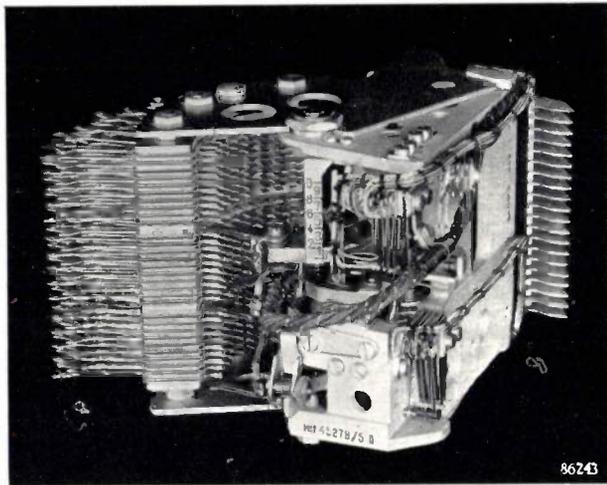


Fig. 2. Two motion selector. The contact bank with  $3 \times 100$  contacts can be seen on the left; in the centre are the three brushes.

circuit. The register can start positioning the selector only after it has wholly or partly received the number that the caller desires. As a consequence of this it is, in general, impossible to avoid some delay between the termination of the last series of dialling pulses and the establishment of the connection, though the selectors used in indirect systems are no slower than those of direct ones.

A second cause of delay in the indirect systems is the fact that in general they are designed as line finder systems, whereas in the direct systems pre-selectors are more usual. The latter have only a small number of contacts whereas line finders have a large number of contacts (see Fig. 1a-b); hence positioning of a line finder will, on an average, take more time than that of a pre-selector.

Although the indirect systems commonly used up till now have had the drawback of a somewhat slower operation as compared with the direct systems, the use of registers nevertheless offers many advantages. These central switching devices are only active while the connection is being established and for that reason they are only engaged for a few seconds during every call. The consequence of this is that it is not necessary to equip all cord circuits with a register of their own: a small number of

registers will suffice for a large number of simultaneous calls. The registers are connected to the cord circuits (*CCt*) via a reducing stage, the cord circuit finder (*CCtF*); see fig. 3. (It is true that this finder is a further cause of delay, because it must first be positioned before the subscriber hears the dialling tone, which is the signal that he may start dialling.) As the required number of registers is small, their complexity has only an insignificant influence on the cost of an automatic exchange. On the other hand there is the great advantage of indirect systems that the provision of special facilities is simpler and less expensive than in the case of direct systems. Some of these facilities will be discussed below.

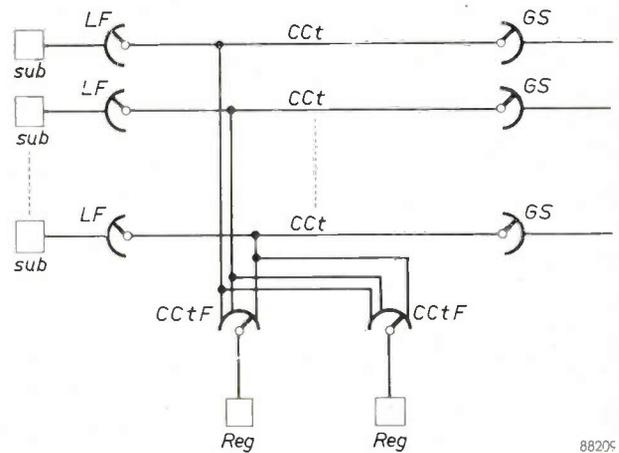


Fig. 3. Connection of the register *Reg* to the cord circuits *CCt* with the cord circuit finders *CCtF* acting as a reducing stage. The other letters correspond to those of fig. 1.

**Outgoing traffic**

If it is required that outgoing calls are established automatically, i.e. without the intermediary of an operator, this imposes certain special requirements on the selectors. In general, it is further required that outgoing calls can be made automatically only from a number of specified extensions. This means that the selectors must then be able to distinguish between those extensions from which outgoing calls may be made automatically and those from which they should pass through the operator. Yet another subdivision of the former group is desirable, viz. the extensions from which trunk calls may be made automatically and those from which these calls should pass through the operator. Finally it may be desirable to block altogether outgoing calls via the operator from certain extensions (normally the operator can connect any extension with an external number by first obtaining the latter via the public

line and then connecting the extension by dialling the internal number).

The many special requirements thus imposed on the selectors and the various circuits imply a certain preference for an indirect system. The demands made by incoming local calls confirm this preference.

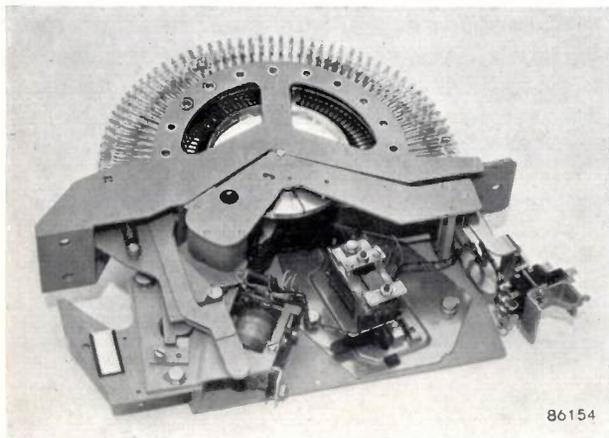


Fig. 4. High-speed uniselector U 45A with four brushes. Other versions exist with more brushes.

To avoid the necessity for bulky cord desks for incoming traffic, modern private automatic branch exchanges are designed so as to be able to handle incoming calls via the automatic equipment. In this way the operator's task, which nowadays consists almost exclusively of answering the incoming calls and dialling the desired internal number, is lightened. However, it is very desirable to have a quicker method for setting up a number than dialling. For this purpose push-button panels have been developed which render it possible to transmit a digit by merely depressing a button. It is not possible to make direct use of the electric signals from the push-buttons for the positioning of the selectors: the signals must first be applied to a register which in turn enables the connection to be established. Some private branch exchanges have been developed that use the direct operation methods for private traffic but push-buttons panels and registers for the incoming local calls. Clearly it is more advantageous, both from a technical and an economical point of view, if registers are used throughout (preferably the same ones being used for both incoming and outgoing traffic). This is the case with the private branch exchange UB 49. Moreover, the development of fast uniselectors and relays have made it possible to speed up the whole register system, the slow response of which was a drawback of earlier indirect systems.

The principal new component is the high-speed uniselector U 45A shown in *fig. 4*. It is a selector with 100 outlets. As it only makes one single rotary movement, its construction is simple and it has the very high speed of 300 contacts per second, as a result of which the selector can hunt all its outlets in only  $\frac{1}{3}$  of a second. It is driven by clutching the rotor of the selector to a continuously-rotating spindle which extends through a number of selectors mounted in parallel. The clutching is controlled by the appropriate register.

The selector is thus not driven step by step, as is the case with the two motion selectors used in direct systems. Here the selector is driven uniformly and as a consequence it runs smoothly in spite of its high speed, which is most important from the point-of-view of wear. Using this high-speed selector reduces the delay between lifting the receiver and hearing the dialling tone to an average of 0.5 sec. Similarly, after the last digit is dialled, the connection is established and the ringing current is then applied within an average time of 0.5 sec. Apart from this obvious advantage, the high speed uniselector possesses other desirable characteristics; these will be discussed later.

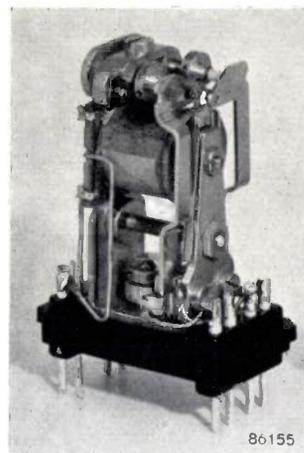


Fig. 5. High-speed relay, type S 50.

A very fast-acting test relay is required to make it possible to stop the selector (running at a speed of 300 contacts per second) stop at any desired contact without error. The relay S 50 shown in *Fig. 5* was developed for this purpose. In an appropriate circuit this relay opens its break contact 0.3 milli-second after the actuating current is switched on. To make possible this extremely high switching speed the relay is of very light construction, yet it is very robust and has a normal contact pressure (20 grams).

### The private automatic branch exchange UB 49

This system may best be described with the aid of a diagram showing the connections of a PABX for more than 100 extensions (*fig. 6*). The various possible connections will now be dealt with in turn.

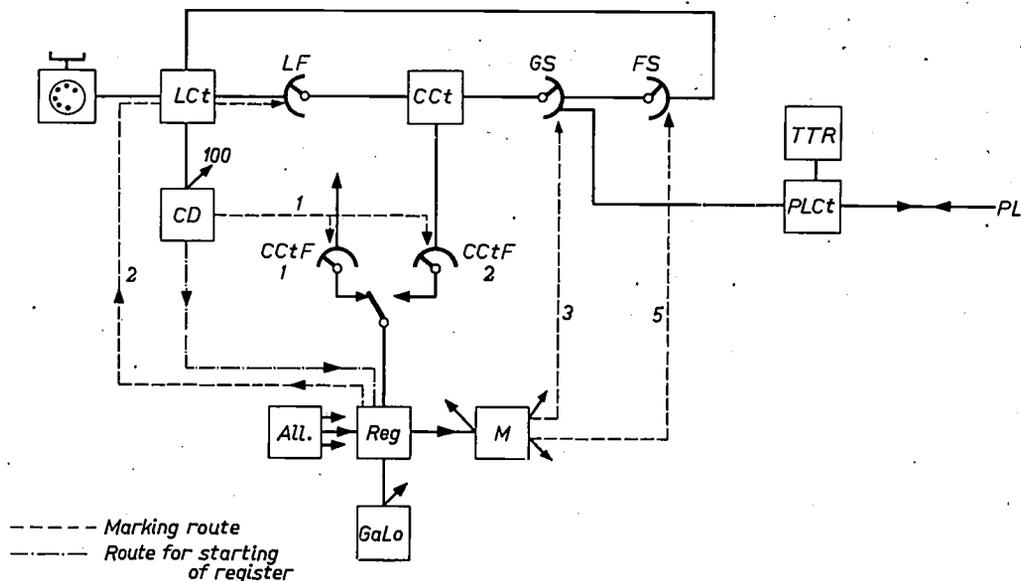
#### Internal connections

For each extension there is a line circuit (*LCt*) which is connected both to appropriate contacts on 15 (or less) line finders and to contacts on the same number of final selectors<sup>3</sup>).

If there are more than 100 cord circuits, then a second cord circuit finder may be introduced as shown in *fig. 6*. It is then possible to connect 200 cord circuits.

When the receiver is lifted from an extension, a signal is sent via the corresponding line circuit (*LCt*) to the call detector (*CD*) which, in turn, applies a starting signal to a register (*Reg*) appropriate to this call. The appropriate register is determined by a call distributor or allotter (*All.*).

As soon as the register in question receives a



88210

Fig. 6. Schematic layout of a PABX type UB 49 for more than 100 connections. *LCt* line circuit for each extension, *LF* line finder, *CCt* cord circuit, *GS* group selector, *FS* final selector, *CD* call detector, *CCtF* cord circuit finder, *All.* call distributor, *Reg.* register, *GaLo* gate lock-out, *M* marker, *PLCt* public line circuit, *TTR* circuit for blocking outgoing automatic trunk calls (trunk traffic restrictor).

Both the line finders (*LF*) and the final selectors (*FS*) are rotary switches with 100 outlets; the extensions are therefore conveniently subdivided into groups of a hundred. A call detector (*CD*), which comes into action as soon as the caller wants to make a call is provided for each group of 100 extensions. A cord circuit (*CCt*) is permanently connected to the brushes of each line finder. This will be explained presently. The cord circuit is also connected to the brushes of a group selector (*GS*) giving access to a number of final selectors and to a number of public line circuits. (*PLCt*). All cord circuits are connected to the contact banks of cord circuit finders (*CCtF*) by means of which a register can be connected with any required cord circuit.

<sup>3</sup>) The selection of this number depends on the number and duration of the calls made by a group of 100 extensions during the busy hour.

starting signal from a call detector, it causes its cord circuit finder to start hunting a free cord circuit whose associated line finder gives access to the group of 100 to which the caller belongs. These free cord circuits are marked as being free by applying a voltage to their contacts in the cord circuit finder via 1. The cord circuit finder is stopped on the contacts of a free cord circuit, by means of a high-speed test relay (type S 50) in the register which is actuated by the marking voltage. When the cord circuit finder has found a free cord circuit, the register starts the line finder of that cord circuit. Now the line relay in the line circuit of the caller's line is activated: hence the caller's line in the contact bank of the line finder is marked by a voltage via 2. This voltage again actuates the test relay in the register, as a result of which the line finder comes to a standstill at the caller's contacts.

As was stated earlier, the extensions are divided into several classes according to whether they are granted the facility of automatic outgoing and/or trunk calls. At the moment the line finder tests the callers' line, the register also determines the class to which the caller belongs. For this purpose class identification wires ("identification multiple") pass along the rows of registers; the appropriate register is connected to these wires at the moment it is designated by the call distributor. In view of the fact that more than one register can never be *simultaneously* active in searching for a caller, it is never possible for more than one register to be connected to the identification multiple at the same time. For every class to be identified there is one multiple-wire and one class relay per register. In every line circuit the contact of the line relay that is to carry the marking voltage to the contact bank of the line finders is connected, by means of a wire, to the multiple-wire of that particular class to which the extension belongs. The caller's line is now marked in the contact bank of the line finder by a voltage from the register via the particular class relay that is connected to the contact of the line relay in the line circuit by means of its multiple-wire. Consequently, both the test relay and the class relay in the register are actuated in the same circuit. As soon as one of the class relays is actuated, the connection between the register and the identification multiple is broken and then the call distributor is in a position to designate another register for answering a call.

In the method of identification described above every possibility of incorrect identification is excluded, as at any given moment only one register is connected to the multiple. This also means that only one register at a time is ever engaged in positioning a cord circuit finder and a line finder. If there are several simultaneous calls, they are not handled at the same time, but consecutively. Such a method can be adopted only if high-speed selectors, such as the Philips uniselector U 45 A, are employed. The cord circuit finder and the line finder are then positioned so rapidly that the delay caused by the consecutive handling of simultaneous calls is negligible.

The positioning of the cord circuit finder and the line finder may be somewhat clarified with the aid of the simplified diagram of fig. 7. This shows the four contact banks of a line finder *LF* and the 8 contact banks of a cord circuit finder *CCtF*; a line circuit, a register and a cord circuit *CCt*, are, however, shown in simplified form.

As soon as the receiver of one of the extensions is removed, the line relay *L* (with three windings) is actuated by the closing of the cradle switch *cs* of the extension. When contact *l*<sub>1</sub> closes the relay *CD* (call detector) is actuated, as a result of

which the register designated by the call distributor (the contacts *r* of this register being closed) attracts the relay *S* via *cd*<sub>1</sub>. This causes the test circuit for the cord circuit finder (with test relay *T*) to be closed in the register by means of contact *s*<sub>1</sub>. Also the release magnet *RM* of the cord circuit finder is actuated by the closing of contact *s*<sub>2</sub>, after which this selector starts rotating. As a result of the closure of contact *cd*<sub>2</sub> all the free cord circuits (*CCt*) relating to the group of 100 lines to which the caller's extension belongs, are marked in arc *I* of the cord circuit finder, for the contact *c*<sub>4</sub> in the free cord circuits is closed, whereas this contact is open when the cord circuit is engaged. As soon as the brushes of the cord circuit finder touch the contacts of a free cord circuit, the high-speed test relay *T* is actuated, as a result of which the release magnet *RM* of the cord circuit finder releases by the opening of contact *t*<sub>1</sub>; the finder then comes to a standstill. At the same time the closing of relay *T* causes the actuation of a relay *A* (not shown in the figure) having contacts *a*<sub>1</sub> and *a*<sub>2</sub>. In this way the test relay *T* is connected into the test circuit for the line finder by way of contact *a*<sub>1</sub>, after which *T* releases and *t*<sub>1</sub> closes again. Consequent on the tripping of contact *a*<sub>2</sub> the release magnet *RM* of the finder is actuated and the latter starts rotating. As soon as the brushes of the line finder are positioned on the contacts of the caller's line, the test relay is actuated again, in the circuit: earth, contact *S*<sub>1</sub>, test relay *T*, contact *a*<sub>1</sub>, contact bank *VI* of the cord circuit finder *CCtF*, contact bank *d* of the line finder *LF*, the closed contact *l*<sub>2</sub> of the line relay, contact *co*<sub>1</sub> of the relay *CO*, the connecting wire to the class-identification multiple, the class-identification wire, contact *r*, the class relay of the class to which the extension in question belongs (*TT* in this case), and the voltage source. Because the test relay is actuated, the release magnet of the line finder *LF* releases on account of the opening of *t*<sub>1</sub>, as a result of which the line finder is stopped. Class relay *TT* is actuated, so that then the class to which the caller's extension belongs is fixed in the register. Even though the subscriber does not belong to the class *LT*, the class relay *LT* is nevertheless energized by the class relay that *does* close (in this case *TT*). Relay *R* releases on account of the opening of contact *lt*<sub>1</sub>, so that the contacts *r* open as well. The register is then no longer connected to the class-identification multiple; nor can it be influenced by the call detector *CD* via relay *S*, if there is another call. As a consequence of the activation of the test relay, a relay *C* (not drawn) in the cord circuit *CCt* is actuated. If the contacts *c*<sub>2</sub> and *c*<sub>3</sub> are then tripped, the result will be that the *a* and the *b* wire of the calling station are connected to the register. The relay *co* is actuated by the closing of contact *c*<sub>1</sub>. The opening of contact *co*<sub>4</sub> disengages the call detector for the next call. The connection between the line circuit and the class-identification multiple is broken by means of contact *co*<sub>1</sub>, while by opening the contacts *co*<sub>2</sub> and *co*<sub>3</sub> the supply of the extension via the line relay *L* is broken.

After positioning of the line finder, the register gives the caller the dialling tone as a sign that the dialling may be started. Three digits will have to be selected, if a layout as shown in fig. 6 is involved. The number of dial impulses of every digit chosen by the caller is counted in the register with the aid of a relay counting circuit. After the counting of a digit the counting relays transfer the result to a group of storage relays, after which the counting relays return to their neutral position. There is no

transfer, however, of the result of the dial impulses of the last digit; in this case the counting circuit also acts as a storage circuit.

After selection of all three digits the register starts positioning the selectors. For this purpose the register must use a common marker, of which there is only one for exchanges up to about 600 numbers.

As soon as the marker is disengaged, the register is connected to it, after which it transfers the dialling data obtained from the caller to the marker. The marker then marks the group of final selectors through which the desired number can be reached in the bank of the group selector and furthermore the line itself is marked in the bank of the final

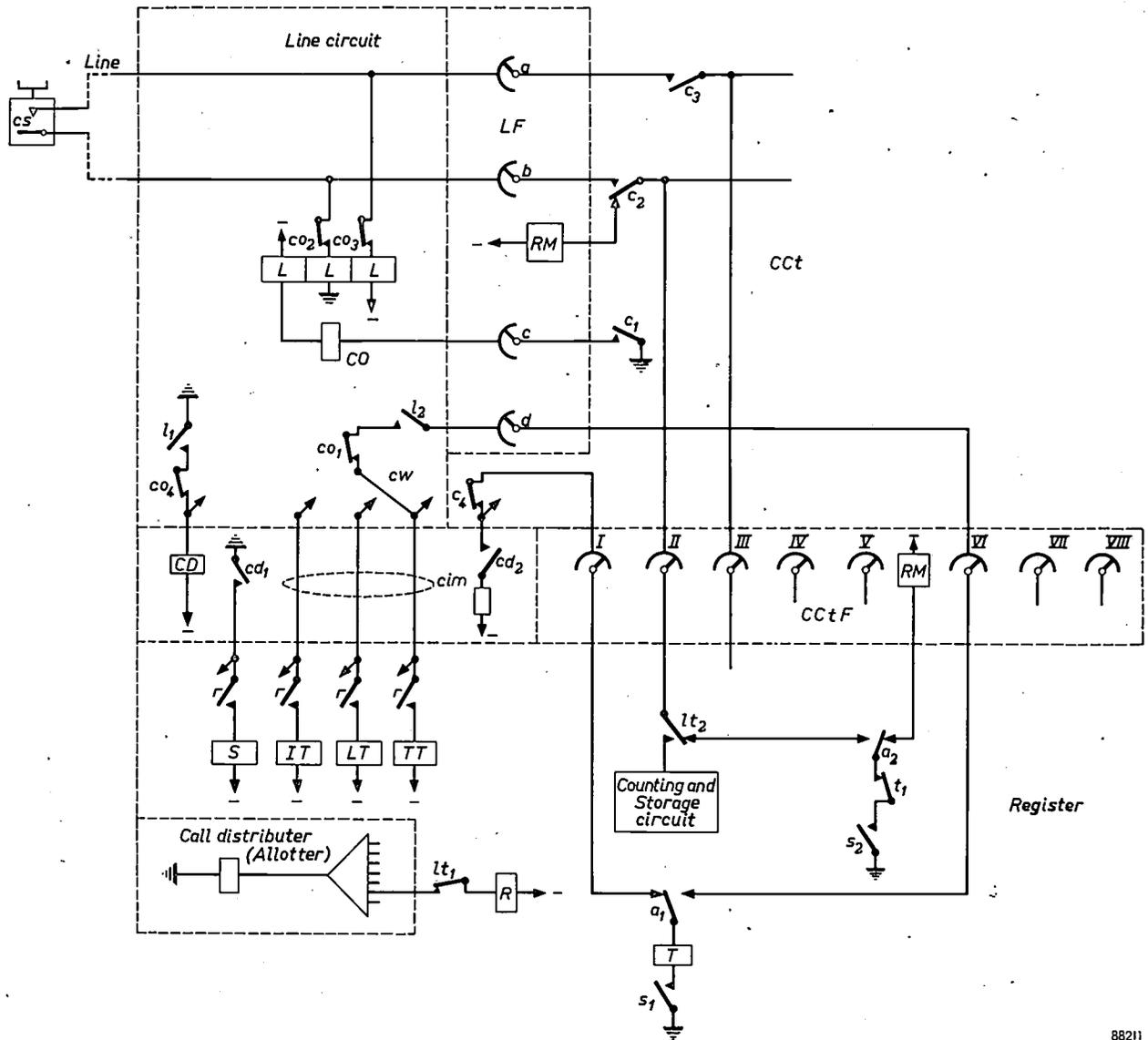


Fig. 7. Simplified diagram of the line finder, the register and the cord circuit. LF line finder with contact banks a, b, c and d; CCt cord circuit; CCtF cord circuit finder with contact banks I, II, etc.; cs cradle switch of the telephone set; CD: call detector; RM release magnets of line finder LF and cord circuit finder CCtF; L line relay; CO cut-off relay; IT, LT, TT: class relay for internal, local and trunk calls respectively; T test relay. The relay contacts are indicated by the same letters (in lower case) as those used for the indication of the relays to which they belong — consequently,  $co_1$  is a contact of relay CO;  $cim$  class-identification multiple;  $cw$  connection from the line circuit to the class-identification multiple. (The function of the contact banks, IV, V, VII and VIII of the cord circuit finder is not discussed here.)

As it is feasible to connect only one register to the marker at a time, there is also a common gate lock-out *GaLo* which ensures that the registers are admitted in groups and dealt with one by one, after which a new group is admitted.

selector (via 3 and 5 respectively in fig. 6). The register consecutively positions the group selector and the final selector, after which the marker is disengaged and available for the next register.

In this case again, only one register out of a group

is enabled to position a group selector - final selector combination. This method offers considerable advantages, as it makes it possible to introduce important simplifications into the circuits and the establishment of double connections, for example, is entirely prevented. This, again, is possible only by the use of the high-speed uniselector U 45A, for then the holding times of the marker are so short that the resulting delays are negligible.

It will be noted that in the entire exchange — if its capacity does not exceed approximately 600 numbers — there are never more than two selectors running at the same time, viz. one in the group selector - final selector category and one in the cord circuit finder - line finder category. As a result of this, the exchange is not only practically silent, but the severe mechanical vibration occurring when many selectors are running simultaneously are also avoided. In this way less electric noise voltages occur and, as a consequence, the system UB 49 is remarkably free from selector noise.

After the positioning of the group selector and the final selector it is checked via the register whether the selected line is engaged or disengaged. If the line to the called station is engaged, then the register breaks the entire connection. All circuits are cleared and the caller hears the engaged signal from his own line circuit. If the called station's line is disengaged, then the register only disengages itself, so as to be in a position to handle further calls. The called station now receives the first ringing current and the subsequent periodical rings from the cord circuit which also supplies the current for the microphones of both the called station and the caller's station.

The microphone current flows via a relay in the cord circuit called the guard relay. This current is interrupted when the microphone is put on the hook; the guard relay then releases, which initiates the process of clearing the connection.

#### *Outgoing traffic via public lines*

Usually the circuits of the system UB 49 are so designed that the lines to the local exchange are accessible when the caller dials one particular digit. These lines are connected to contacts in the banks of the group selector, and the register designated by the call distributor starts positioning it as soon as the pulses from the aforesaid single digit are received. When discussing internal connections it was mentioned that the register has determined the class to which the caller's extension belongs before the dialling tone is emitted. If he is not allowed to make outside calls, i.e. if his line relay is not

connected to the relevant wire of the identification multiple, then the register does not start the group selector, but it breaks the entire connection built up so far and the caller hears the engaged signal from his line circuit. If he is entitled to make local calls, then the register adjusts the group selector to a free public line circuit. Furthermore it applies a special signal to the cord circuit and disengages itself for the next call. The result of the special signal is that the supply in the cord circuit is not switched on and that the speech wires are connected from the line circuit to the public line circuit *PLCt* (fig. 6), from which the microphone is now fed via a guard relay in the public line circuit. The question whether the connection is broken or not, thus no longer depends on the cord circuit, but on the public line circuit.

The group selector is positioned to a free public line with the aid of the marker by means of a marking connection from the register to the contact bank of the group selector, via the marker and the free public line circuits. This connection also serves for passing the signals to the public line circuit in order to indicate whether or not the caller is entitled to make trunk line calls, a datum that is fixed in the register by the actuation of the relevant class relay (*IT*, *LT* or *TT*, see fig. 7). If the caller is not authorized, then this signal switches on a device in the public line circuit which prevents trunk calls from being made (*TTR*). The way in which such a device operates depends on the way in which trunk connections are established in a given country. We shall not enter further into this matter here.



Fig. 8. Operators desk for handling the local traffic of a PABX type UB 49 with an arbitrary number of public lines.

The exchange UB 49 enables a caller who has built up an outgoing public line connection to consult with another person in the private system during this call. The way in which this is done will be discussed in the next section.

#### *Incoming local calls*

In the case of incoming calls on the public lines the operator's assistance is needed for the connection of these lines to the extensions. For this purpose use must be made of a control desk (*fig. 8*) or a control set (*fig. 9*). In principle both installations

- e) If the called person's line is engaged, then the operator may interfere in the conversation and offer the outside call. If the called person does not wish to interrupt the conversation at once, then the operator has to try again after some time.
- f) The called person answers and the conversation proceeds.
- g) The called person requests the operator to transfer the call to another extension.
- h) The called person may return call to the operator without any comment.

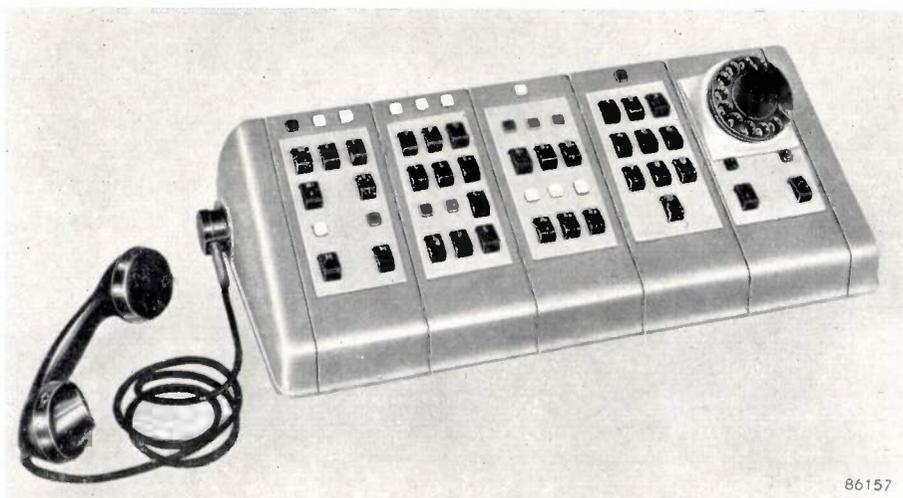


Fig. 9. Operator's set having the same function as the desk in fig. 8.

offer the same facilities. The (smaller) control set will mainly be used in those cases where the operator has not got a fulltime job; in such cases it is more economical to put the operators set on a normal desk at which the operator does her other work.

A new system is used for the standard version of desk and set, which will now be described. In this system the public lines no longer appear individually on the operator's desk. In order to estimate the new technique at its true value, it is necessary to consider the various possible situations which can arise during a call from outside.

- a) The call has been received, but has not yet been answered.
- b) The call has been answered, but has not yet been put through.
- c) The desired person is not immediately available after being called. In this case the public line is switched to the waiting position and remains engaged while the operator handles other calls in the meantime.
- d) If there is no reply from the extension called, the operator has to inform the caller of this fact.

- i) The caller may request to be consecutively connected to various extensions. In this case the operator must see to it, by carrying out a special switching operation, that the call is passed back to her desk, whenever the receiver of an extension is put down.
- j) The operator must receive a special signal, if she has chosen an extension which may not receive local calls, or if she has carried out some incorrect manipulation.

When all the public lines are permanently connected with the operator's desk, it is not possible to fit more than two or three lamps per line for lack of room. One should then try to indicate the various situations mentioned above with the aid of lenses of different colours, lamps dimmed or full on, or with lamps that flicker at various frequencies. The picture of the lamp panel, however, soon becomes distracting in that case, while the signalling system still does not come up to standard.

For the above reasons another solution has been chosen. One lamp and an associated push button are now provided for each of the possible situations. The

public lines are no longer in permanent connection with the operators desk. This connection is established only for one line at the moment at which the push button is operated and only the situation of that particular line is indicated on the desk. Moreover, there are lamps for showing that one or more connections to the public exchange are in one of the situations *a, b, c, d, e, g, h* or *i*, and that the operator must take some action. Another advantage of this system is that the design of the operators desk is independent of the number of public lines to be controlled. If there is too much local traffic for one operator, then two positions can simply be connected in parallel. The signal lamps for all the situations enumerated above, for which the operator's assistance is called in (*a, d, e, g, h* and *i*) are white. The waiting lamp is red, as the situation indicated in this way requires the operator's special attention. Green-coloured lamps indicate whether the operator's desk is connected to a public line or to an extension.

the appropriate button, the public line finder *PLF* of the corresponding operator's circuit *OpCt* is started, as a result of which she is connected to the caller's exchange line. Once the caller has told her to whom he wishes to speak, she depresses the button for internal interconnection and she is then connected to the transfer line of the public line circuit (*TrL*), which is connected to the banks of the line finders as a normal line to an extension, the operator makes a normal call, so that a register is started. This register hunts for a free cord circuit and positions the latter's line finder to the transfer line. The operator is thus connected to a register in the same way as an extension and she receives the dialling tone from the register.

At the moment when the register positions the line finder to the transfer line (*TrL*) it is also ascertained that the call comes from an operator. The means by which this is done will not be discussed here, as this would carry us too far. As a consequence

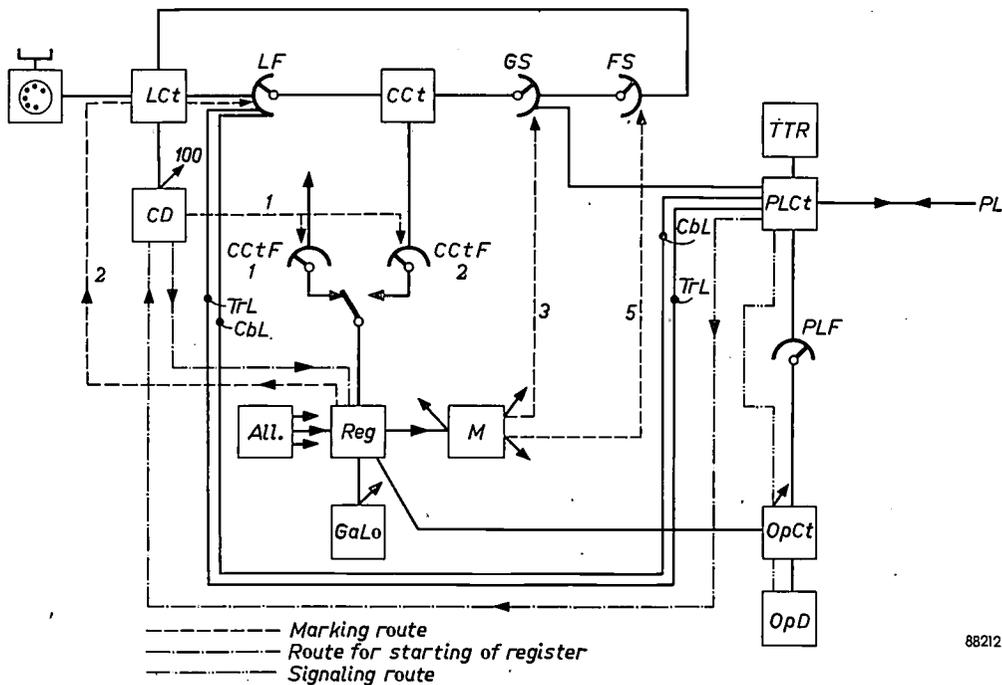


Fig. 10. Layout of a PABX type UB 49 for more than 100 extensions (see also fig. 6), with provisions for the operator and for call-back added to it. *PLF* public line finder, *OpCt* operator's position circuit, *OpD* operator's desk, *TrL* transfer line, *CbL* call-back line. For other letters see fig. 6.

88212

We shall now give a brief description of the way in which an incoming local call is put through (fig. 10). If there is an incoming call on one of the public lines *PL*, a general calling lamp lights up on the operators desk *OpD*. If the operator depresses

a direct connection is established via 4 wires between the operator's position circuit and the register. The four-wire connection enables the operator to store the digits in the register very rapidly by means of the push buttons on her desk. The

connection is established within 0.5 sec after the button is depressed for the last digit, as high-speed selectors are incorporated.

In this case as well, the cord circuit is connected through while being controlled by the register. The latter then ascertains by means of a subsequent test whether or not the caller is entitled to receive local calls. If not, then the private connection is broken and a lamp on the desk warns the operator. If, however, the caller is authorized to receive local calls, then his extension is fed from the public line circuit. whence the call is made. In that case the operator may either remain in the circuit until the called person answers, or withdraw from it at once by depressing a button provided for the purpose. In the latter case the called person's station is automatically connected to the public line, if he answers the call. If he does not, however, the call is automatically passed back to the operator's desk after about 20 sec. If the called person's line is engaged, then the operator can enter into the circuit and offer the local call. By depressing a button provided for the purpose the operator can re-establish the connection with the calling subscriber in order to inform him of her findings.

Once the connection between the public line and the extension has been established, the called person using the extension may consult with another person using an extension without breaking the connection with the public line. In order to achieve this, the public line circuit is connected to the contact bank of the line finders by another line, the call-back line (*CbL*). As soon as the button for calling back on the extension is depressed, a call starting from the public line circuit is made by way of the call-back line. The public line is temporarily cut out, but retained in the public line circuit by means of an auxiliary circuit. At the same time the station connected to the transfer line is connected with a cord circuit and a register by way of the call-back line. As soon as the dialling tone is heard, the desired number is dialled, after which both extensions are connected in the usual way. If the button for call-back of the first extension is then depressed again, the connection via the call-back line is broken and the connection with the public line is re-established. If, however, the button for call-back of the second extension is depressed, then the connection between the public line circuit and the first extension — by way of the transfer line — is broken and the second extension is connected to the public line by the call-back line. This procedure can be repeated as often as desired; the transfer line and the call-back line then change their function alternately.

### Overflow group selectors

Once the exchange has attained a certain size it is no longer possible to connect all final selectors and public lines to every group selector. In most cases the group selectors are then divided into two (or, in large exchanges, into more) parts, after which the final selectors of each group of a hundred extensions are uniformly distributed over these parts. The consequence is, however, that the final selectors are accessible to a limited extent only. For, if a certain group selector is chosen, then only part of the final selectors in the hundred-group required are accessible. All these final selectors may be engaged, while it may be possible that the final selectors for that particular hundred-group which are connected to the group selector of the other part are still disengaged. In that case the efficiency of the final selectors and exchange lines drops <sup>4)</sup>.

In an exchange with more than about 600 lines the system UB 49 obviates this disadvantage by using overflow group selectors. A certain number of the final selectors for every group of one hundred extensions (e.g. 7 or 8 out of 15) are directly connected to every group selector (*GS*, fig. 11).

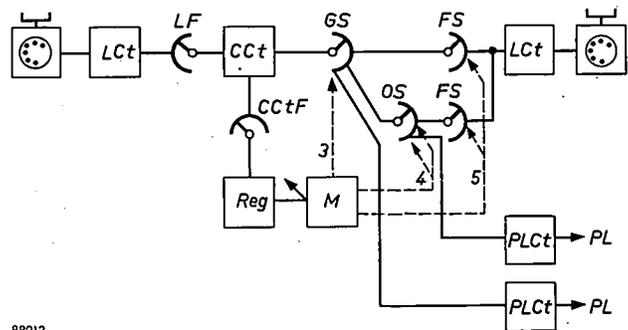


Fig. 11. Diagram of connections with overflow group selectors. *LCt* line circuit; *LF* line finder, *CCt* cord circuit; *GS* group selector; *OS* overflow group selector; *FS* final selector; *CCtF* cord circuit finder; *M* marker; *PLCt* public line circuit; *PL* public line.

The remaining final selectors are connected to the contacts of each of the overflow group selectors (*OS*). It is then checked for all calls whether the final selectors connected to the group selectors are disengaged. The portion of calls that cannot be handled by these final selectors is passed on to the

<sup>4)</sup> By the efficiency of, say, a group of lines we understand the ratio of the traffic (number of call-minutes per hour) that can be handled by this group with a certain grade of service, to the traffic that would occur if all lines were continuously engaged. The efficiency of a large group is higher than that of a small group, if the grade of service remains the same. This means that a group of, say, 100 lines can handle more traffic than two groups of 50 lines each, with the same grade of service. See for example J. Atkinson, *Telephony* (footnote <sup>2)</sup>.

overflow group selectors. These selectors thus receive the remaining calls from all groups of one hundred stations; the overflow group selectors still get a fairly large number of calls to deal with and consequently their efficiency is good.

The marking (via the lines 3 and 5, if a direct connection can be established, and via the lines 3, 4 and 5, if the connection is established by way of an overflow group selector) is ensured by the marker circuit (*M*) in combination with the register. In the case of an indirect connection the marker circuit brings the register to the state required for the positioning of three selectors, viz. *GS*, *OS* and *FS*.

The local traffic is, in a similar way, offered to the public line circuits (*PLCi*) connected to the group selectors and then to the circuits connected to the overflow group selectors.

### Networks

If various scattered private exchanges have to co-operate (external private traffic), a cable network must be laid between these exchanges. The system UB 49 can be adapted to any required network configuration. The use of a central marking device makes it possible to determine the most favourable route for every connection in this external traffic. The marking device also investigates whether the traffic can be passed to the other exchange directly, or — if all lines to this exchange are engaged — by way of a third exchange (overflow traffic). In mesh-shaped networks a considerable improvement of the efficiency of the groups can be achieved in this way.

### Special provisions in the system UB 49

Although they do not belong to the normal equipment of the UB 49 exchange, all kinds of special devices can be fitted;

a) *Night traffic installation.* Normal equipment may

control the public lines during the evening and night hours and on Sundays in the operator's absence. After a normal extension has answered a local call the desired extension may be called by way of the call-back line; this extension may then take over the call.

b) *Conference device.* A conference device enables a maximum number of 10 users of extensions to hold a telephonic conference, either together or in groups. The connections are controlled by the operator who first dials all the desired extensions, after which the bells of all these sets are rung simultaneously.

c) *Automatic staff location.* Persons who are wanted on the phone can be warned by means of acoustic and optical signals. The device in question is switched on by dialling a call number, which causes signals to be automatically given according to a fixed code. As soon as the called person dials a certain number (the report-number) on an arbitrary extension, he is automatically connected to the caller's extension.

The provision of all these facilities presents no special difficulties and need not be described further here.

---

**Summary.** After an introduction dealing with the general development of private telephone exchanges, a description is given of the UB 49 exchange. This operates on the register system in which the characteristic disadvantage (slower action) of register systems is overcome by the use of a high-speed uniselector, the Philips U 45 A. The characteristic advantages, (reduction of equipment in large exchanges), however, are retained, while many additional advantages are obtained. For instance, the system requires a minimum of maintenance and it is remarkably free from selector noise.

A special operator's desk has been developed for the establishment of incoming calls. The special feature is that the public lines are not permanently connected to the desk; they are connected to it only at the moment when control is required. This renders the design of the desk independent of the number of public lines to be controlled. In addition much more complete signalling of the various situations applying to the public lines is possible. The introduction of selection by means of push-buttons makes it possible to pass on local calls to extensions very quickly.

## ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk \* can be obtained free of charge upon application to the Philips Research laboratory, Eindhoven, Netherlands.

**2269:** S. Duinker: A shifting register using ferro-resonant flip-flops (Appl. sci. Res. B4, 317-328, 1955, No. 5).

Ferro-resonant effects, i.e. instabilities occurring in the steady state response of non-linear circuits by gradual variation of some circuit parameter are briefly discussed and the application of these effects in trigger and flip-flop circuits is pointed out. A shifting register using ferro-resonant flip-flops is described in which the information is shifted by unidirectional pulses. By simple means it is possible to shift the information in both directions depending upon the polarity of the shifting pulses. Some details of an experimental register for a telex-on-radio system working at a supply frequency of 20 kc/s and permitting shifting pulses at a repetition rate of 0.8 kc/s are discussed. Attention is paid to the design considerations especially for the application of supply frequencies in the Mc/s range.

**2270:** J. J. Opstelten and N. Warmoltz: A double-sided micromanometer (Appl. sci. Res. B4, 329-336, 1955, No. 5).

A condenser plate is mounted close to each side of a thin diaphragm, separating two chambers. The two capacities thus formed are part of a bridge circuit which is fed by a high frequency oscillator. An amplifier with a narrow bandwidth, a rectifier and a micro-ammeter are used as a null-indicator. At one side of the diaphragm a pressure prevails, much lower than the pressure to be measured; at the other side the gas is admitted the pressure of which is to be measured. The displacement of the membrane by the gas pressure is compensated electrostatically by the aid of a calibrated potentiometer. The reading of this potentiometer is a linear measure for the pressure as a result of the differential way of compensating. The range is from  $10^{-5}$  mm — 1 mm of mercury pressure difference at any absolute value. The apparatus is made of chemically fairly resistant materials, and its indication is independent of the nature of the gas, so that it can be used for almost every gas.

**2271:** J. Bloem and F. A. Kröger: A relation between hardness and stoichiometry in lead sulphide single crystals (Nature 175, 861-862, 1955).

Single crystals of PbS were prepared and reheated under various sulphur pressures at 1200 °K to control the stoichiometric composition. After rapid cooling of each crystal to room temperature, the concentration of free charge carriers (electrons or holes) and the hardness were determined. It was shown that stoichiometric crystals contained the lowest carrier concentrations. Reheated at higher sulphur pressures, an excess of sulphur (consistent with lead ion vacancies) was introduced together with positive holes. At low sulphur pressures sulphur ion vacancies together with free electrons were present. These concentrations vary from about  $10^{-3}\%$  to  $10^{-1}\%$  when reheated under various sulphur pressures. The hardness of the crystals also showed a minimum value for stoichiometric crystals; for strong *n* and *p* type crystals the highest values were observed.

**2272:** W. J. M. Vestjens: Een mannetje van *Dixippus morosus* Br. (Entomologische Berichten 15, 404-406, 1955). (A male of *Dixippus morosus* Br; in Dutch).

*Dixippus morosus* has been bred since 1950, and from 1952 onwards, on a larger scale in cages about  $20 \times 20 \times 25''$ . In August 1954 a male occurred. It is distinguished by its smaller size, by a red stripe at the ventral side between coxae I and II, and by the lack of the red marking at the inner side of the anterior femurs.

**2273:** L. A. A. Sluyterman: Molecular weight of insulin as derived from paper electrophoresis (Biochim. et Biophysica Acta 17, 169-176, 1955).

Some uncertainty exists as to whether the molecular weight of insulin is 6000 or 12000. The experiments described in this paper may help to decide this question. Samples of insulin were treated with various amounts of acetic anhydride under conditions favourable for specific acetylation of the amino groups. Thus products with various contents of free amino groups were obtained. Paper electrophoresis of these products demonstrated the progress of the acetylation reaction by the successive appearance of new bands and the disappearance of the initial insuline band. A total number of three new bands were found. This shows that acetylation

of insulin occurs in three steps, which suggests the presence of three amino groups in one molecule of insulin. According to chemical analysis insulin contains three amino groups in each unit of molecular weight 6000 (Sanger, 1945). Hence the present experiments indicate that the molecular weight of insulin is 6000.

**2274:** J. H. Stuy: Photoreactivation of ultraviolet-inactivated bacilli (*Biochim. et Biophysica Acta* 17, 206-211, 1955).

The phenomenon of photoreactivation has been studied with three cultures of bacilli. It has been shown that they could be reactivated by light with wavelengths of 3655, 4047 or 4358 Å after ultraviolet inactivation. Since these radiations also have an inactivating effect, the photoreactivation may be small, even not noticeable, when the dose-rate of the incident light is too great. Bacterial spores did not show photoreactivation.

**2275:** E. W. van Heuven: On the noise of fluorescent lighting installations (*Acustica* 5, 101-111, 1955).

In the frequency spectrum of the noise of a fluorescent lamp ballast, "hum" and "rustle" can be distinguished. Hum comprises small even multiples of the mains frequency, at which the dimensions of the ballast are small in comparison with the wavelength of the sound in the air. The noise at these frequencies is, therefore, mainly produced by the ballast acting as a vibration exciter to larger mounting surfaces. Rustle frequencies (about 1000—3000 c/s) are radiated directly by the ballast itself. Hum may be eliminated by use of a core construction where the displacements due to magnetic attraction across the air-gap cancel out the magnetostrictive displacements at the mounting points of the reactor. Rustle radiation must be isolated by putting the ballast in a closed box. A filling for the empty space in these boxes providing acoustical impact sound insulation and good thermal conductivity is described.

**2276:** J. Bloem, F. A. Kröger and H. J. Vink: The physical chemistry of lead sulphide in relation to its semiconducting properties (Rep. Bristol Conf. Defects in Crystalline Solids, 1954, Phys. Soc. London 1955, p. 273).

The semiconducting properties of lead sulphide single crystals have been measured after heating at various temperatures in controlled atmospheres. Thermoelectric power measurements indicate an energy gap of 0.3 eV at room temperature, and an effective electron and hole mass of about 0.25 of the

electron mass in free space. With the use of the theory of Kröger, Vink and van den Boomgaard it was possible to deduce the width of the energy gap and the position of the various vacancy levels as a function of temperature, from 800 to 1200 °K.

**2277:** F. van der Maesen and J. A. Brenkman: On the behaviour of rapidly diffusing acceptors in germanium (*J. Electrochem. Soc.* 102, 229-234, 1955, No. 5).

The acceptor activity of Cu and Ni and their diffusion in germanium have been investigated. Their action as acceptors as well as recombination centres gives reason to believe that impurity atoms are placed substitutionally in the lattice. However, the value of  $10^{-5}$  cm<sup>2</sup>/sec of the diffusion constant makes it likely that the diffusion goes interstitially. This leads to the concept of an equilibrium existing between substitutional and interstitial atoms. The fact that the diffusivity is dependent on the range of diffusion can be explained qualitatively on the basis of the picture given above,

**2278:** H. C. Hamaker: De betekenis van de statistiek voor de ontwikkeling van de experimentele wetenschap (*Sigma*, 1, 55-58, 1955, No. 3). (The significance of statistics in the development of experimental research; in Dutch).

The application of statistics to technical problems is shown to be a basically new phase in the development of technology.

**2279:** F. K. Lotgering: Order-disorder phenomena in iron sulphides Fe<sub>1-x</sub>S (*Z. Phys. Chemie* 4, 238-241, 1955 No. 3-4).

Iron and sulphur form a continuous series of compounds with compositions varying from FeS to Fe<sub>7</sub>S<sub>8</sub>. The sulphides with a low sulphur content are antiferromagnetic, those with a high sulphur content are ferromagnetic while an anomalous ferromagnetic behaviour appears at compositions between these two regions. The magnetic and crystallographic properties are experimentally investigated and explained using the simple theory of order-disorder phenomena.

**2280:** T. Holmes: Something new in sodium lamps (*Public Lighting* 20, 662-664, 1955).

Description of a sodium lamp constructed of a glass with improved resistance to sodium vapour. The use of this glass is made feasible by employing a xenon-neon mixture in place of the argon-neon filling commonly used in the past to reduce the ignition voltage.

# Philips Technical Review

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

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

## PHASE LINEARITY OF TELEVISION RECEIVERS

by A. van WEEL.

621.375.121.018.782.3:621.397.62

The picture signal that controls the beam current of the picture tube in a television receiver must be a faithful copy, both in amplitude and in phase, of the signal generated in the television camera. The properties thus required of the vision channel of receivers could be met fairly simply while there were comparatively few television transmitters in existence, i.e. while receivers were not required to be particularly selective. The situation has changed, however, since transmitters became more numerous, and the selectivity of receivers had to be improved accordingly. In this connection, difficulties have arisen with regard to the phase linearity. In some quarters recourse has been made to compensating defective phase linearity at the transmitting end. The author of this article puts forward another solution, that of a phase-linear I.F. amplifier, which turns out to be no more expensive than that of conventional I.F. amplifiers.

It should be noted that the work described in this article is based upon the Gerber standard (so named after the Chairman of the C.C.I.R. sub-committee which proposed this standard) now in use in a number of countries. This standard prescribes negative picture modulation, and frequency modulation for the sound. In these respects it is similar to the American standard, but different from the British standard. As a consequence, the author's conclusions are directly applicable to the American system, but only with certain restrictions to the British system.

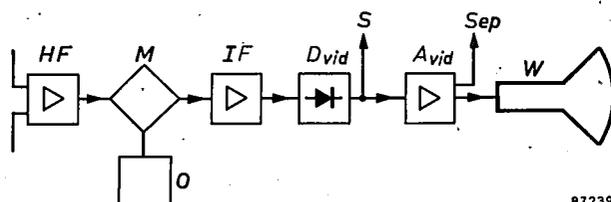
The "vision channel" of a television receiver embraces all the circuits from the aerial terminals to the picture tube. The aerial delivers a carrier that is amplitude-modulated with the picture signal. The vision channel has the task of producing a signal exactly corresponding to that modulation for application to the control electrode of the picture tube. A block diagram of the vision channel appears in fig. 1.

The picture signal is a complicated function of time, being proportional to the brightness values of the various picture elements scanned one after the

other in the camera. The networks of the vision channel must be designed so as to reproduce this time function undistorted.

It is possible to compute the shape of the output signal from any circuit, given the shape of the input signal. A *step function* (fig. 2a) is frequently chosen as input signal; the amount of distortion exhibited by the output signal (fig. 2b) gives a good idea of the effect of the network on the picture signal it passes<sup>1)</sup>.

In practical television engineering, however, it is not usual to employ this method for individual sections of the circuit. In preference, use is made of Fourier analysis whereby any time function can be regarded as the sum of a large number of sine functions. If the time function is not periodic the Fourier spectrum consists, not of a number of discrete lines (as does that for a periodic time function), but of a continuous band. Thus, with a non-periodic time function, the power contained in the original is distributed over a certain range of frequency.



87239

Fig. 1. Block diagram of the vision channel in a television receiver. HF high frequency amplifier. O local oscillator. M mixer stage. IF intermediate frequency amplifier. D<sub>vid</sub> video detector. A<sub>vid</sub> video amplifier. W picture tube. Sep sync. separation stage. In the intercarrier sound system, the sound signal is taken off at S.

<sup>1)</sup> Cf. J. Haantjes, Philips tech. Rev. 6, 193-201, 1941.

The reason for preferring this method of approach to that of step functions is that it is usually easier to calculate how a network transfers single sinusoidal components at different frequencies than to investigate directly how the network in question would react on the time function itself.

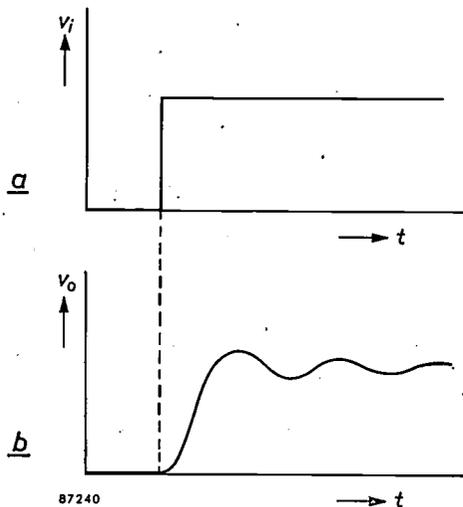


Fig. 2. a) Step function and (b) typical step function response.  $v_i$ : input voltage, and  $v_o$ : output voltage of a network, plotted as function of time  $t$ .

Knowing the frequency range within which the Fourier components of the signal fall, the networks are required to pass this frequency range without distortion. This means that the intensity relationship between the various components may not be changed (a change of strength of all components in the same ratio is of course permissible). Likewise, the individual components may not undergo differing delays (again, all components may be delayed to the same extent, since this merely means delaying the signal as a whole). Furthermore, no new components such as noise, hum and non-linear distortion products may be added; in the present article, however, we shall not consider these matters.

What is required, then, of all sections of the circuit along which the signal has to pass is that both their amplitude characteristics and their delay-time characteristics shall be flat for the frequency range in question.

A sinusoidal component of the input signal may be written  $\hat{v} \sin \omega t$ ; at the output this becomes  $a \hat{v} \sin \omega(t - \tau_f)$ . The quantity  $a$ , plotted as a function of frequency, gives the *amplitude characteristic*; the quantity  $\tau_f$  similarly plotted gives what is called the *phase-delay characteristic*<sup>2)</sup>. The latter represents the retardation or delay undergone by a given phase of the component in question. The

retardation may also be interpreted as the phase angle  $\varphi = \omega \tau_f$  that arises on account of the network. To make the phase delay independent of frequency, the phase-angle must increase linearly with frequency. It follows from this that for the undistorted transmission of a signal all the networks employed must be phase-linear.

#### Relationship between the amplitude and delay characteristics

It is worth while to consider the causes of the delay in signal transmission just discussed. In line transmission it is easy to see that the finite velocity of propagation of the waves along the cable involves a certain delay. In networks consisting of coils, capacitors and resistors the physical distance to be travelled by the signal is very short; that marked delays nevertheless occur in them is a consequence of the fact that a certain time is necessary for both the building up of an electric field in a capacitor and the building up of a magnetic field in a coil. Only in a perfect resistor are the voltage and the current exactly in phase.

It is impossible to build amplifying circuits that, apart from tubes, contain only ohmic resistances, for resistors necessarily involve a certain amount of stray capacitance, and so do tubes. The effect of these capacitances can be reduced by employing lower resistance values, but then the amplification falls off as well.

It is possible to obtain high and constant amplification within the desired frequency band with the aid of special networks ("wide-band circuits"). The wider the band and the higher the amplification required, the greater is the number of coils and capacitors needed in these networks. The inertia of the network therefore also increases. In general, this additional delay will not be the same for all frequencies, so that the delay characteristic will vary all the more as the effectiveness of the network is improved from the point of view of amplitude alone. If on the other hand the network is so designed that the delay properties are improved, then there is a deterioration in the amplitude response.

The demand for as much amplification as possible within a certain frequency band thus creates difficulties with regard to the amplitude characteristic or the delay characteristic or to both; over and above this, the need to satisfy certain selectivity requirements outside the desired band also leads to undesired effects within the band. Thus, under the Gerber Television Standard (CCIR), the video channel must exhibit an amplitude characteristic which is flat up to about 5 Mc/s and which falls to 5% at

<sup>2)</sup> H. J. de Boer and A. van Weel, An instrument for measuring group delay, Philips tech. Rev. 15, 307-316, 1953/54.

5.5 Mc/s (under this standard the frequency difference between sound and picture carriers is 5.5 Mc/s). All that is required of the delay characteristic is that it should be flat up to 5 Mc/s; the delay of suppressed components is of no importance in television receivers.

The amplitude characteristic described above can always be obtained with the aid of sufficiently complicated filters. To make the amplitude characteristic run flat up to a given cut-off frequency and then fall very steeply, certain resonance effects are made to occur in the neighbourhood of the cut-off frequency. Now, resonance phenomena have the property of taking a certain time to come into action; passive resonant networks can naturally supply no energy of themselves and it is only possible to obtain high amplitudes of oscillation by a building-up process. The inference is that the phase delay of such networks is always markedly greater in the region of the cut-off frequency than in the remaining part of the frequency band.

In general, minimum phase-shift networks are employed for this purpose. These are networks which at all frequencies have a smaller phase-shift than that obtainable with other networks giving the required amplitude characteristic. The question of whether a given network is of minimum phase-shift type or not, may be resolved by the possibility or otherwise of drawing an equivalent circuit of the network in the form of a cascade of  $\pi$  or T sections, or both, using only realizable circuit elements (see fig. 3a). If this is impossible on account of bridging elements (fig. 3b), then the network is, in general, not one of a minimum phase-shift type.

It must be borne in mind that the criterion may only be applied once the network has been reduced to its simplest form, should this be necessary. If, for example,  $Z_1$ ,  $Z_2$ ,  $Z_3$  in fig. 3b are impedances of the same kind, a  $\pi$ -T transformation makes it possible to reduce the whole circuit to a single T-section made up of realizable elements (fig. 3c). Hence the network is in fact of a minimum phase-shift type.

In the practical design of filters that have to exhibit a linear phase characteristic within the bandpass range, a common procedure is first to give the amplitude characteristic the desired shape with the aid of minimum phase-shift networks, and then to add a special kind of non-minimum phase-shift network that comes within the category of "all-pass" networks. These have the property of combining a horizontal amplitude characteristic for all frequencies from zero to infinity with a phase delay which in one way or another varies with frequency; they thus offer an opportunity of compensating the phase errors of the first filter.

Such circuits are accordingly to be found in professional television equipment, e.g. in radio relay systems, in amplifiers for coaxial cables and, as will be discussed later, in certain television transmitters.

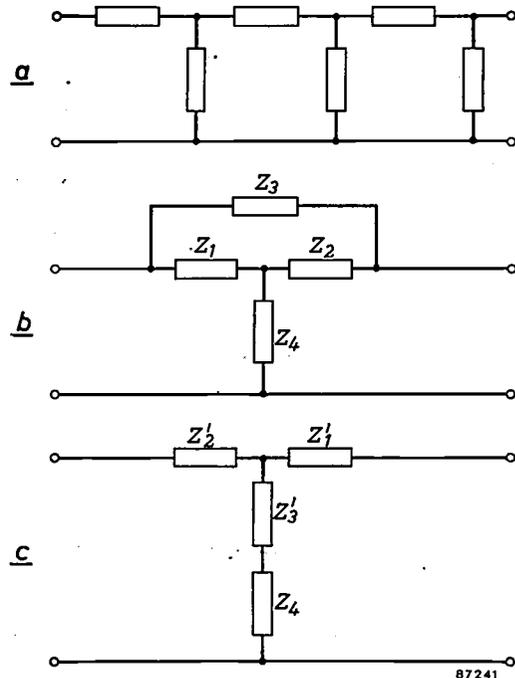


Fig. 3. a) Example of a network consisting of a T-section and  $\pi$ -section in cascade. If the network is made up entirely of realizable circuit elements, then it is of a minimum phase-shift type.

b) If  $Z_1$ ,  $Z_2$  and  $Z_3$  are impedances not of the same kind, the network is not, in general, of minimum phase-shift type. If they are of the same kind, the delta arrangement can be turned into a star arrangement, as in (c), so that network (b) is in fact of minimum phase-shift type (provided that all the impedances are realizable ones).

The use of such phase-compensating networks in normal television receivers is impossible for economic reasons. However, experience has shown convincingly that there is too much variation in the phase delays in modern receivers. These errors are much more troublesome in recent sets than they were in older ones — a direct consequence of the growth in the number of transmitters, which has made it necessary to increase receiver selectivity. The increased selectivity has been achieved by the employment of circuits that fulfil that purpose but at the same time involve greater phase-delay errors.

The selectivity of a television receiver, like that of a normal broadcast receiver, is obtained in the IF amplifier. The general rule that higher selectivity means bigger discrepancies in the phase delays is valid here as elsewhere. In the IF amplifier there is amplification of an IF carrier modulated with the video signal. The absolute value of the delay under-

gone by the carrier is in no way important for the quality of the picture received; the necessary condition is that the carrier and the side-band components should all have the same delay. As was explained in the article referred to above<sup>2)</sup>, it is not necessary in such a case that the phase delay  $\varphi/\omega$  of the IF amplifier should be constant over the pass-band; it is sufficient that the *group delay*  $\tau_g = d\varphi/d\omega$  should satisfy this requirement because, when it does so, the phase delay of the modulation is the same for all video frequencies.

The Gerber Standard recommends an amplitude characteristic for the receiver as well as one for the transmitter (fig. 4). It will be seen from fig. 4b

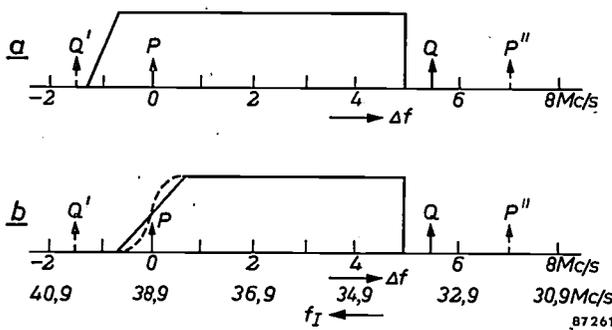


Fig. 4. a) Frequency spectrum of a transmitter according to the Gerber Standard of the C.C.I.R. (tolerances omitted). The horizontal scale represents difference of frequency  $\Delta f$  from the picture carrier. P indicates the picture carrier and Q the sound carrier. P'' is the picture carrier of the next channel above, and Q' the sound carrier of the next channel below. b) The corresponding amplitude characteristic for the receiver. The scale of intermediate frequencies  $f_I$  is applicable to the (usual) choice of 38.9 Mc/s as IF for the picture carrier.

that at the receiving end only one side-band has to be dealt with in respect of modulation frequencies of more than 0.75 Mc/s. The phase-delay  $\tau_{fm}$  of the modulation in this single side-band is given by:

$$\tau_{fm} = \frac{1}{\omega_m} \int_{\omega_0}^{\omega_0 + \omega_m} \tau_g d\omega, \dots \dots (1)$$

where  $\omega_0$  is the angular frequency of the carrier and  $\omega_m$  the angular frequency of the modulation.

For modulation frequencies under 0.75 Mc/s there is asymmetrical double side-band reception. Here the phase delay of the modulation follows, in a somewhat more complicated fashion, from the ratio of the amplitudes of the two side-bands as well as from the magnitude of  $\tau_{fm}$  as given by equation (1) for each side-band separately (the limits of integration for the lower side-band being  $\omega_0$  and  $\omega_0 - \omega_m$ ; see also the article referred to in<sup>2)</sup>). Here too it remains true that, if the group delay is constant within the IF band to be amplified, then the phase

delay of the modulation is constant for all video frequencies; on the other hand, deviations from a flat group delay characteristic generally give rise to discrepancies in the phase delay of the modulation frequencies.

The IF amplifier must therefore possess a flat group delay characteristic over the pass-band. The amplitude characteristic within this range — and outside it too (for reasons of selectivity) — must fulfil certain requirements. It is not necessary for the amplitude characteristic of the receiver to embody a slanting straight flank as shown in fig. 4b or an approximation to it. Any characteristic that is centrally symmetrical about the carrier frequency (like the dotted curve in fig. 4b) provides a flat amplitude curve for the modulation.

### Effect of the amplitude characteristic on the shape of the signal

At the highest modulation frequencies the amplitude characteristic is rectangular in shape (fig. 4b). It is not *a priori* certain that cutting off the picture frequency band gives the best quality picture. To explain this in more detail, let us consider an imaginary network having an exactly rectangular amplitude characteristic that cuts off abruptly at the frequency  $f_{lim} = \omega_{lim}/2\pi$  (fig. 5a), and giving a constant delay-time for all frequencies; for the sake of simplicity we shall suppose the delay-time to be zero. If now such a network is fed with a signal having the form of a step function (fig. 5b) it can be fairly easily shown that the output signal will be given by the integral sine function:

$$v_o(t) = \frac{2}{\pi} \int_0^{\omega_{lim}} \frac{\sin \omega t}{\omega} d\omega. \dots (2)$$

The shape of this function of  $t$  can be worked out by expanding it as a series; fig. 5c shows the result. The following three properties of the function will be noted:

- (1) The output voltage deviates from zero even before  $t = 0$  (and even when  $t = -\infty$ );
- (2)  $v_o$  is not infinitely steep at  $t = 0$ ;
- (3) On both sides of  $t = 0$  the curve goes beyond the initial and final values; it oscillates about them and approaches them as the distance from  $t = 0$  increases.

That the function should "begin" for negative values of  $t$ , i.e. even before a voltage is present on the input, is of course physically impossible. The result obtained is a consequence of the assumption that the delay-time of a filter of this kind is zero, and this too is physically unobtainable. A filter

with an amplitude characteristic like fig. 5a can be approached as closely as desired if one is prepared to extend it with more and more sections. A perfectly rectangular amplitude characteristic will

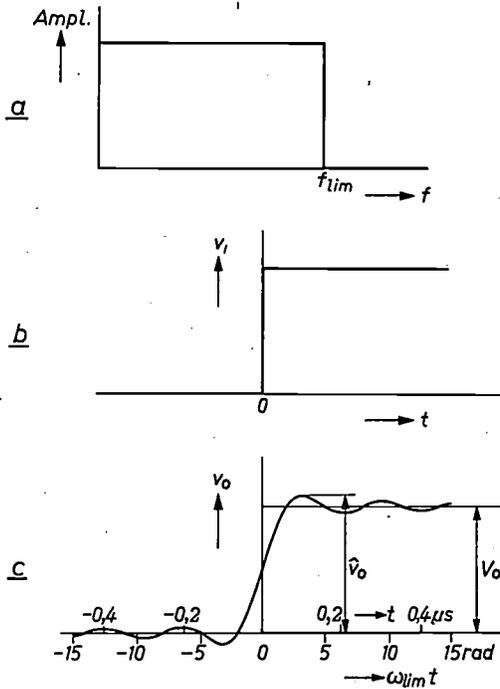


Fig. 5. a) Flat amplitude characteristic of an imaginary network, falling abruptly to zero at the cut-off frequency  $f_{lim}$ . b) Step function (cf. fig. 2a). c) Response to (b) of network with amplitude characteristic as in (a). The overshoot factor is defined as  $(\hat{v}_0 - V_0)/V_0$ .

only be obtained when an infinite number of sections is employed. The filter will then have an infinitely long delay-time; in other words, the instant  $t = -\infty$  at the output according to fig. 5c is the same as the instant  $t = 0$  at the input according to fig. 5b. This explains the apparent irreality of point (1) above.

It may be added that it is possible to arrive at a fairly close approximation of the amplitude characteristic in fig. 5a using only a limited number of elements. In such a case, phase errors will occur and these will exactly compensate the oscillations in fig. 5c during the times prior to zero; here too, therefore, no output signal will arise before the input signal is applied.

The second property mentioned above, the finite slope at  $t = 0$ , is directly connected with the bandwidth available. The abscissa in fig. 5c is given in units of  $\omega_{lim}t$ ; the greater  $\omega_{lim}t$ , the steeper will be the rise in the curve when plotted as a function of  $t$ . Thus the bandwidth determines the steepness obtainable.

The third property, the occurrence of "undershoot" and "overshoot" — as the fluctuations around the initial and final values may be termed — is directly connected with the amplitude characteristic's abrupt fall to zero at the cut-off frequency. It can be shown<sup>3)</sup> that these oscillations do not occur when the amplitude characteristic, instead of being rectangular, is Gaussian in form, i.e.

$$A = A_0 \exp(-a\omega^2), \dots (3)$$

as in fig. 6a, curve I. If the parameter  $a$  in (3) is chosen such that the Gaussian curve has already dropped considerably when the cut-off frequency  $f_{lim}$  of fig. 5a is reached, it will be seen that the reproduction of the step function will be much less steep (see fig. 6b). A suitable compromise between steepness and amount of undershoot and overshoot is obtained by giving the amplitude characteristic the form of a Gaussian curve up to the cut-off frequency, where it will have fallen to one half, and then letting it drop to zero abruptly (curve II in fig. 6a). The initial oscillation then has an overshoot factor of about 5%, as against 9% for the integral sine function.

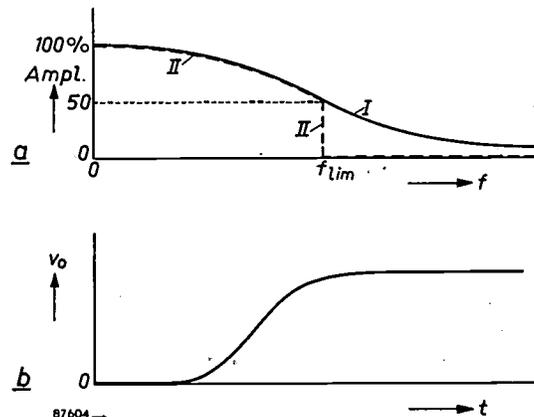


Fig. 6. a) I amplitude characteristic of Gaussian form. II the same, except that the characteristic falls abruptly from 50% to zero at the cut-off frequency  $f_{lim}$ . b) Step function response of a network having the amplitude characteristic I in (a).

### Selectivity requirements and phase errors

So far we have only considered the amplitude characteristic for the modulation signal. We shall now lay down in greater detail the selectivity required outside the admitted band. Firstly, the sound of the desired channel must be suppressed (the carrier frequency of the sound signal is separated from the picture carrier frequency by 5.5 Mc/s);

<sup>3)</sup> See, for example, C. Cherry, Pulses and transients in communication circuits, Dover Publications, New York 1950, p. 175, or K. Küpfmüller, Die Systemtheorie der elektrischen Nachrichtenübertragung, S. Hirzel, Zurich 1952, p. 51.

secondly, the risk of interference from transmitters adjacent in frequency must be considered. Since a band of 7 Mc/s is available for each television channel the picture carrier is separated from the sound carrier of an adjacent channel by only 1.5 Mc/s ( $Q'-P$  in fig. 4). The sound signal from a transmitter working on the adjacent channel may cause interference when the receiver is tuned to a weak transmitter. In these circumstances the signal from the adjacent sound carrier penetrates as far as the video detector and sets up a video frequency component of 1.5 Mc/s, which can produce very troublesome effects in the picture.

For these reasons it is necessary that the neighbouring sound carrier should be heavily attenuated; in good receivers the standard is that it should be 1% or less of the maximum of the IF amplitude curve (100-fold attenuation). On the other side of the pass-band, the sound carrier of the desired channel has to be suppressed. This implies 200-fold attenuation, but it need not be effected entirely in the IF amplifier; the corresponding video frequency (5.5 Mc/s) lies outside the picture frequency band, and the sound of the desired channel can therefore be suppressed partly in the video amplifier. In receivers using the intercarrier sound system<sup>4)</sup>, 20-fold attenuation (to 5%) is often laid down for the IF section.

None of this is particularly difficult to put into practice as long as only the amplitude characteristic need be considered. But complications proceed from the effect mentioned above, whereby discrepancies in delay-time arise near frequencies corresponding to abrupt changes in the amplitude characteristic. This is illustrated by the curves in figs. 7a and b; the former is the selectivity curve<sup>5)</sup> and the latter the group delay curve of the IF amplifier of a normal television receiver. If a comparison is made between the curves from the middle of the band outwards, it will be seen that on both sides the group delay starts to change before the selectivity. It seems as if coming events with regard to the amplitude response cast their shadows before them in the group delay behaviour.

The effect of group delay changes in the IF section on the modulation signal is much more serious at the picture carrier end (i.e. for low modulation frequencies) than it is at the other end of the admitted band (high modulation frequencies). This may be inferred from a calculation of the phase

delays of different modulation components by means of equation (1). In this equation the group delay is integrated over a certain frequency band and then divided by the modulation frequency in question. Two hypothetical group delay curves are drawn

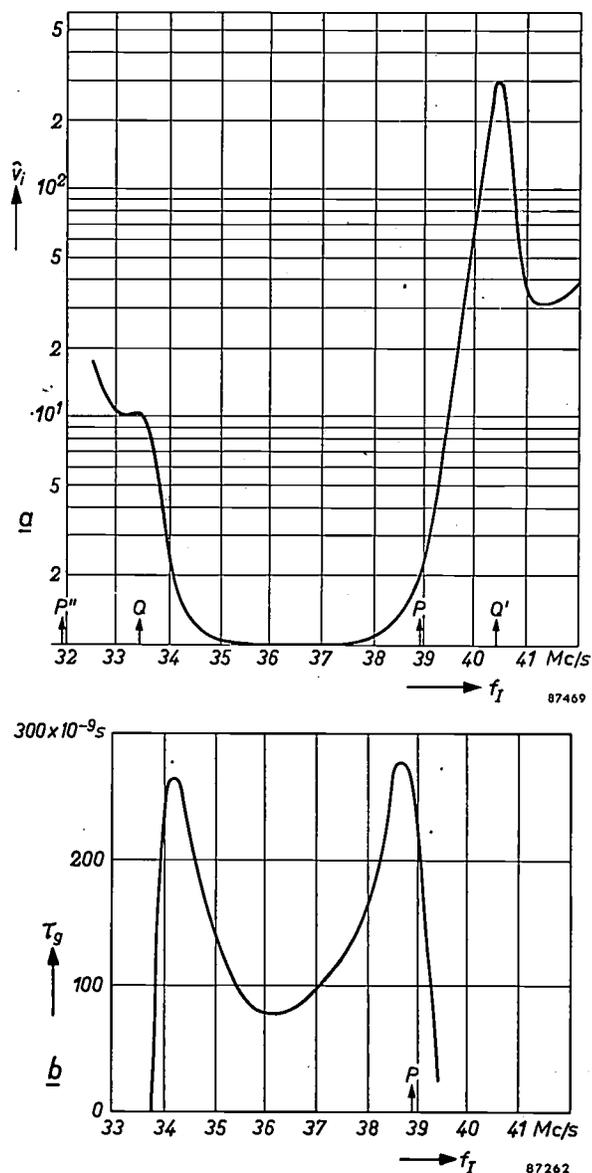


Fig. 7. a) Selectivity curve and (b) group delay characteristic of the IF amplifier of a normal television receiver. (In (a) the input signal  $\hat{v}_i$  required to give a certain output signal is plotted on a logarithmic scale.) With regard to P, Q, P' and Q', see fig. 4.

in fig. 8a: curve I exhibits a change in  $\tau_g$  in the vicinity of the carrier frequency but is otherwise flat; curve II exhibits a change of the same magnitude, but at a point well away from the carrier frequency. It will be seen that the integral of equation (1) has the same value for  $\omega_m = \omega_{m1}$  in curve I as for  $\omega_m = \omega_{m2} \gg \omega_{m1}$  in curve II (the two hatched areas in the diagram are equal); but to find the values of  $\tau_{fm}$  we still have to divide

<sup>4)</sup> See for example Philips tech. Rev. 15, 198, 1954/55.

<sup>5)</sup> The selectivity curve is the curve obtained by plotting as ordinates the reciprocals of the ordinates of the amplitude curve.

the integral by  $\omega_{m1}$  and  $\omega_{m2}$  respectively. Since  $\omega_{m1}$  is smaller than  $\omega_{m2}$ , the group delay characteristic represented by curve *I* has a greater interfering effect on the picture than that represented by curve *II*.

In fig. 8*b* the same effect is demonstrated in the group delay curve measured on an actual IF amplifier. In this case the carrier frequency was 38.9 Mc/s. The phase delay ( $\tau_f$ ) of the video frequency signal (insofar as it originates in this side-band) was determined by graphical integration from the group delay characteristic as measured. It may be seen that the big changes in  $\tau_g$  at high video frequencies cause only limited changes in the phase delay.

Thus the phase errors arising from the incorporation of trap circuits for the adjacent sound channel — which, as indicated by fig. 4, lies only 1.5 Mc/s away — are much more troublesome than the phase

errors that occur at higher modulation frequencies (up to 5 Mc/s) owing, for example, to the suppression of the desired sound carrier (which lies 5.5 Mc/s away). In practice, the phase errors of the first-mentioned, more troublesome, variety manifest themselves mainly as a smudging of abrupt black-to-white transitions in a left-to-right direction.

Fig. 9 illustrates the control voltage response at the picture tube for two step functions, one a step

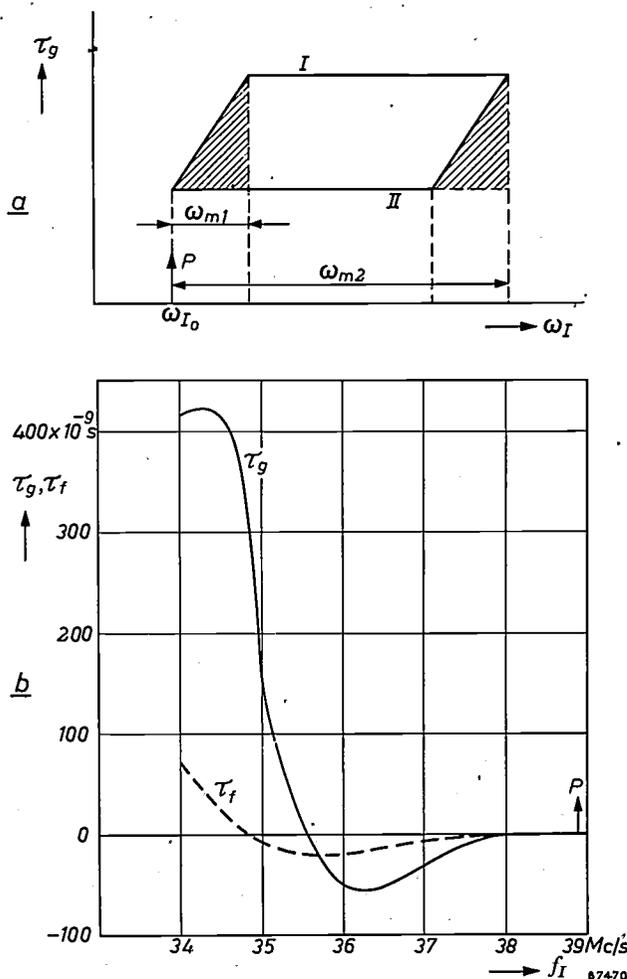


Fig. 8. a) Comparison of two hypothetical group delay characteristics (*I* and *II*) for an IF amplifier. *II* is better than *I*, since its rising portion is further away from *P*, the IF of the carrier.  
 b) Measured group delay characteristic  $\tau_g$  of a certain IF amplifier. The phase delay time  $\tau_f$  has been derived from  $\tau_g$  by graphical integration.

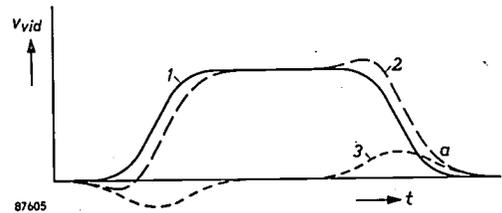


Fig. 9. Response to a white-to-black transition, followed by that to a black-to-white transition.  $v_{vid}$  control voltage on the picture tube. The full black line *1* represents the case where phase distortion is absent, the broken line *2* (which is the sum of curves *1* and *3*) the case where it is present. The divergence of *2* from *1* at *a* (on the extreme right) manifests itself in the picture as a grey smear.

from white to black and the other a step from black to white. The full line represents the case in which the phase distortion under discussion is absent, the broken line a case where it is present. The bulge into blacker-than-black preceding the second step is, of course, not visible on the screen; it can be seen, however, that the jump in brightness is not from black to white but from black to grey producing a gradual, smudged, transition towards white. The received picture suffers a considerable loss of definition in consequence.

**Possible solutions**

There are three ways in which the unfavourable effects of phase errors might be removed:

- (1) by compensation in the video section of the receiver,
- (2) by compensation in the video section of the transmitter, and
- (3) by the use in the receiver of an inherently phase-linear IF amplifier.

The last-named method seems at first to be the most obvious, but up to now it has not been used. This is a consequence of a curious situation which has developed in the field of television receiver design. It is generally known that phase properties are of importance in television; nevertheless, the phase characteristic as such has not hitherto been taken into consideration, measured or checked either in the design or the manufacture of television sets. This applies to almost all the monochrome receivers at present in existence.

Two reasons can be advanced to explain this strange state of affairs. For one thing, the measurement of phase delay characteristics has become a subject of study only since the war. Various equipments for this purpose have meanwhile been developed, but in general they have been so complicated as to be not suitable for employment in development laboratories, to say nothing of employment in the factory. It is only quite recently, at the Philips Research Laboratories in Eindhoven, that apparatus

in the early days of a new technical development.

It may be added that one kind of check on phase properties is always possible, viz. the judgement of picture quality; further, an impression of the receiver's behaviour in this respect may be gained by displaying on an oscillograph the response of the whole receiver to a step function. In these two ways it is quite feasible to decide whether distortion is present, but to pronounce on the kind and the cause of the distortion is another matter.

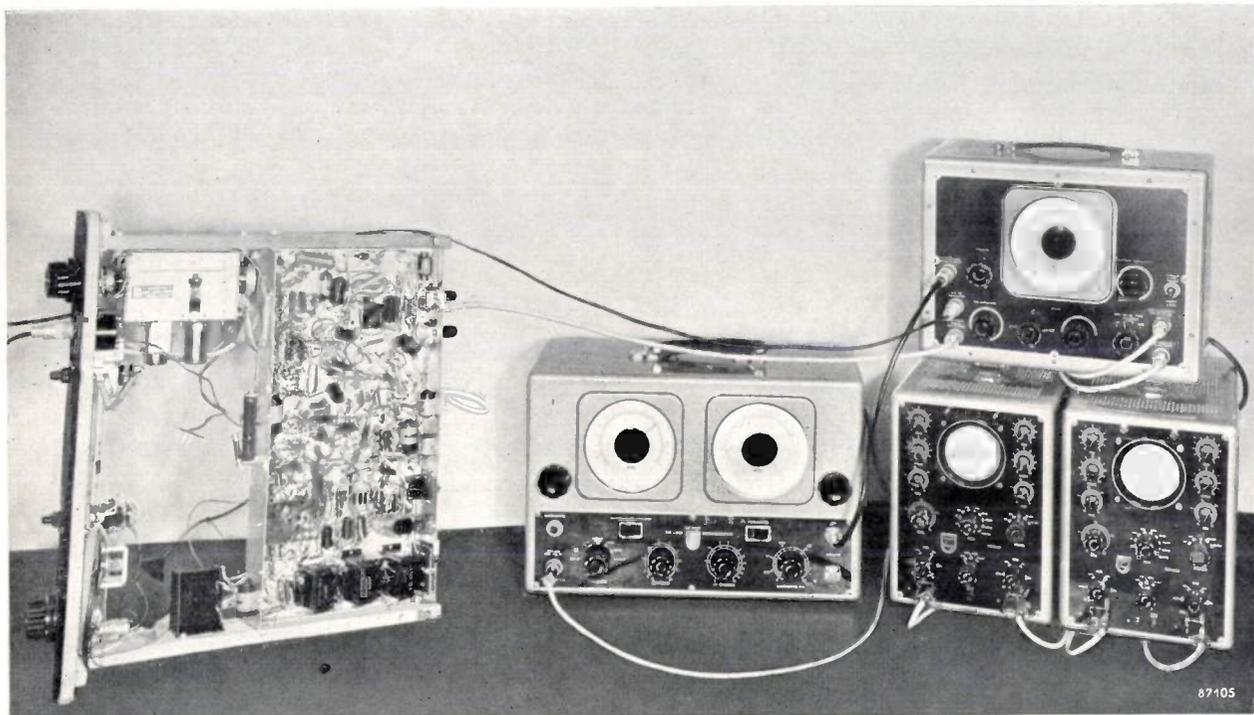


Fig. 10. Checking the selectivity and group delay characteristics of the IF amplifier in a television receiver.

*Left:* Receiver chassis. *Centre:* Wobbulator (Type GM 2889 signal generator). *Right:* Two oscillographs and, resting upon them, a device by means of which the selectivity curve and the group delay characteristic are displayed simultaneously on the left-hand and right-hand oscillographs respectively. This group delay meter has been in production for some time, under the type code GM 2894.

(*fig. 10*) has been developed<sup>6)</sup> which allows the group delay characteristic to be displayed directly on the screen of a cathode-ray tube, and which is so simple that it can be used for checking in manufacture.

Thus the direct measurement of phase properties was a difficult matter until recently. The design of television receivers was therefore necessarily based on the amplitude characteristic alone. In the beginning results were fairly good. This can be attributed to the fact that in those days it was not necessary to make the standard of selectivity as high as it is today. Apart from this, it is, in general, not usual to aim extremely high

In the meantime the falling-off in picture quality due to improved selectivity was not watched idly in development laboratories. Means to overcome it were sought in an empirical fashion. Giving the amplitude characteristic of the VF amplifier a pronounced hump between 1 and 2 Mc/s — until then, of course, it had been made as flat as possible — seemed to provide a remedy, up to a point. It did indeed effect a considerable improvement in picture quality. Thorough investigation of the processes associated with this modification of the VF amplitude characteristic has shown that it created phase errors in the video amplifier that almost cancelled out those in the IF amplifier. Amplitude distortion was also involved, it is true, but its effects

<sup>6)</sup> C. J. Heuvelman and A. van Weel, *Wireless Engr.* **33**, 107-113, May 1956.

proved to be less troublesome than those of the original phase errors. Thus picture quality was improved, without however becoming optimum for the given bandwidth.

Given the carrier frequency and the amplitude and the group delay characteristics of an IF amplifier, it is possible to calculate accurately the phase distortion undergone by the video frequency signal. The phase errors can then, in principle, be compensated to any desired extent in the video frequency section with the aid of the all-pass networks mentioned above. This is not done in practice, for it would make the receiver too expensive. It is therefore a case of making do with the the imperfect but inexpensive solution described above.

A further solution — point (2) above — is to use video frequency phase compensation at the transmitter. At the transmitting end the question of cost is of much less importance, and in no way prohibits the use of this method. One disadvantage that suggests itself is that, for this solution to be universally acceptable, all television receivers would have to have the same IF group delay characteristic; this is of course not so in practice. There are other disadvantages associated with this solution; we shall return to them at the end of the article.

It is perhaps useful to devote a few words to the phase properties of the transmitter. The filter that has to suppress the unwanted side-band in the antenna signal likewise causes considerable phase errors; it is not practicable to compensate them in the radio-frequency section on account of the high power that has to be fed to the antenna. Consequently the compensation is always carried out in the video frequency section of the transmitter, where it must balance the phase errors due to the vestigial filter as well as possible.

The effect of the compensation is checked by observing the picture appearing on the screen of a receiver, say, of a standard commercial type. If the compensation is adjusted in such a way that picture quality is optimum, then the phase errors of the receiver being used are compensated at the same time. All those who possess a set of this type will therefore be able to receive a good quality picture, but the quality of picture given by receivers with better phase properties will be worse!

For this reason, compensation at the transmitter of phase faults in the receiver must be based on a standardized receiver delay-time curve. The Nordwestdeutsche Rundfunk has worked out this problem systematically. The basis chosen was a receiver IF amplitude characteristic which fulfilled certain demands as to selectivity and which was capable of realisation with the aid of minimum phase-shift networks. On this basis it is possible to work out the corresponding group delay curve, and upon this the compensation in all transmitters and the design of future receivers has to be based. This solution, as adopted by the Nordwestdeutsche Rundfunk, is doubtless one of the best thought-out practical forms of the system of transmitter compensation.

The reasons that recourse was made to phase compensation at the transmitting end was the assumption that a selective receiver could only be made phase-linear by the employment of complicated and therefore expensive circuits. In what follows we shall show that this belief is unjustified and the need for phase compensation at the transmitter is then seen in quite a different light.

#### Requirements to be fulfilled by a phase-linear IF amplifier

Before going into the ways and means of realizing a phase-linear IF amplifier, we shall set down the requirements to be fulfilled as regards the amplitude characteristic.

*Inside* the pass-band, the amplitude characteristic must fall by a factor of two at the IF carrier frequency, and elsewhere possess the desired shape from the standpoint of the video frequency modulation.

*Outside* the pass-band it must satisfy the following demands:

- (1) At the frequency of the sound carrier, the amplification must fall to approximately 5% (it may not be lower with the system of intercarrier sound<sup>4</sup>) in general use in the CCIR system).
- (2) At the frequency corresponding to the sound carrier of the *adjacent* channel the amplification must fall to 1% or less.
- (3) Beyond this frequency the suppression may be less severe, since the modulation depth of the highest picture frequencies in the neighbouring picture channel is always fairly slight. Under the Gerber Standard (C.C.I.R) the ratio of the powers of sound and picture carrier is 1 : 5. The sound carrier may be regarded as a single side-band modulation of the picture carrier with a frequency of 5.5 Mc/s. With 100% depth of modulation, the power of each side-band component would be a quarter of the carrier power. With a sound power/picture power ratio of 1 : 5, as above, the depth of modulation of the sound carrier, regarded as side-band component, is  $\sqrt{4/5} \times 100 = 89\%$ . If it is enough to attenuate this "modulation component" to 1%, then an attenuation to 5% will suffice for the much weaker picture-modulation components of the highest frequencies.
- (4) For the two adjacent picture carriers, lying at +7 and -7 Mc/s from the desired picture carrier, the signal must likewise be attenuated to 1% or less; this can always be achieved without difficulty.

Information as to how these demands on the amplitude curve are to be met and how at the same time a linear phase characteristic within the pass-band may be obtained, is provided by the fundamen-

tal relationship between the phase and amplitude characteristics of networks with minimum phase shift. In an earlier section it was mentioned that any change in the shape of the amplitude characteristic necessarily affects the shape of the phase or delay-time characteristic. For networks with minimum phase-shift it is possible to give an analytical relationship connecting the amplitude and phase characteristics such that the one curve is wholly determined when the other is known for all frequencies from zero to infinity. This relationship, known as Bode's relation <sup>7)</sup>, can be written as follows:

$$\tau_{fc} = \frac{\omega_c}{\varphi_c} = \frac{2}{\pi} \int_0^\infty \frac{A - A_c}{\omega^2 - \omega_c^2} d\omega, \dots (4)$$

where  $\tau_{fc}$  is the phase delay at the angular frequency  $\omega_c$ ,  $\varphi_c$  the phase shift at this same frequency and  $A$  the "insertion loss" (a logarithmic measure of the attenuation in the network concerned, i.e. the amplification in the network is proportional to  $e^{-A}$ ).

For the time being we shall disregard the ideal amplitude characteristic and attempt to work out from equation (4) what sort of amplitude characteristic results from the requirement that the phase characteristic should be linear. The frequency band for immediate consideration will be that of the video frequencies, that is, from zero to a certain cut-off frequency.

For the phase characteristic to be linear, the phase delay must be constant; that is, the value of the integral in (4) must not depend on  $\omega_c$ . This is so if  $A$  is a quadratic function of  $\omega$ :

$$A = a\omega^2, \dots (5)$$

since the quantity to be integrated then reduces to  $a$ . However, this gives an infinite delay. It is not necessary to seek a physical explanation for this, for the application of equation (4) to an amplitude characteristic in accordance with (5) is not permissible, the reason being that certain limitations must be introduced in deriving (4) and these limitations exclude a curve in accordance with (5) <sup>8)</sup>.

Nevertheless, it is possible to use the quadratic-exponential amplitude curve given by (5) as a basis for further investigation, owing to the effect of the weighting factor  $1/(\omega^2 - \omega_c^2)$  in the integral

of (4). This weighting factor is very large in the immediate vicinity of  $\omega_c$ . This being so, the phase angle and phase delay at these frequencies — although of course depending on the values possessed by  $A$  throughout the whole frequency range from zero to infinity — will be determined largely by the way that  $A$  changes in the neighbourhood of  $\omega_c$ .

This property allows the following statement to be formulated: *if it is ensured that the attenuation factor  $A$  is a quadratic curve up to a certain cut-off frequency and that outside this frequency band it does not diverge too rapidly from this quadratic form, then the phase characteristic may be expected to be linear within the said frequency band, with significant deviations occurring only in the vicinity of the cut-off frequency.* This expectation is confirmed in practice, as we shall see later.

Thus our first conclusion is that the amplitude characteristic of a phase-linear video frequency amplifier within the pass-band is given by  $\exp(-a\omega^2)$ , that is, that the characteristic is a Gaussian curve.

However, we were concerned not with video frequency amplifiers but with IF amplifiers, i.e. with band-pass networks. Now there is a network theorem which says that any video frequency characteristic can be practically realised in the form of half the characteristic of a bandpass filter (see fig. 11) providing that the band it passes is relatively narrow.

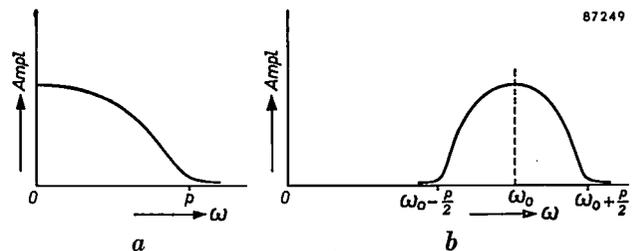


Fig. 11. a) Amplitude characteristic of a low-pass network over the range  $\omega = 0$  to  $\omega = p$ . b) Amplitude characteristic of a band-pass network over the range  $\omega = \omega_0 - \frac{1}{2}p$  to  $\omega = \omega_0 + \frac{1}{2}p$ . Curve (a) can be regarded as the right-hand half of curve (b) which has been shifted through a distance of  $\omega_0$  to the left, the abscissa being expanded by a factor of 2.

By virtue of this principle, the amplitude and phase properties discussed above can be transferred directly to the band passed by a filter. Thus, to give a bandpass filter a linear phase characteristic, the selectivity characteristic must take the form of a Gaussian curve within the pass-band. IF selectivity curves are usually drawn with a logarithmic amplitude scale; the Gaussian curve is then a parabola.

<sup>7)</sup> H. W. Bode, Network analysis and feedback amplifier design, Van Nostrand, New York 1953, pp. 302-336.  
<sup>8)</sup> J. Peters, Einschwingvorgänge, Gegenkopplung, Stabilität; J. Springer, Berlin 1954, pp. 43-44.

Before going into the question of whether a curve of this kind is suitable for the transmission of a picture signal modulating a carrier wave, we shall show how far a Gaussian amplitude curve is compatible in practice with a linear phase characteristic or, in this case, with a flat group delay characteristic.

In wide-band amplification it is usual to work with staggered circuits; this consists of employing LC elements that are slightly detuned and differentially damped with respect to one another in the various stages of a cascade amplifier. The LC elements are so designed that the amplitude characteristic of the whole amplifier is as flat as possible.

It is also possible to design the elements so that the group delay is as nearly constant as possible. In figs 12 and 13 amplitude and group delay curves for one half only of the pass-band are drawn for an amplifier containing various numbers of LC elements. In fig. 12 the aim was constant amplitude, in fig. 13 constant group delay. It will be seen from fig. 12 that, as the number of circuits  $n$  increases, the amplitude curve remains flat up to an ever higher cut-off frequency, in the vicinity of which, however, the group delay curve deviates more and more from flatness. In fig. 13 it is seen that, the greater the number of LC elements, the greater is the range of frequency over which there is an almost constant group delay, while the amplitude characteristic within this range exhibits ever greater divergences.

In order to show to what degree the amplitude curves of fig. 13 approximate to a Gaussian function, we have transferred them to the logarithmic scale

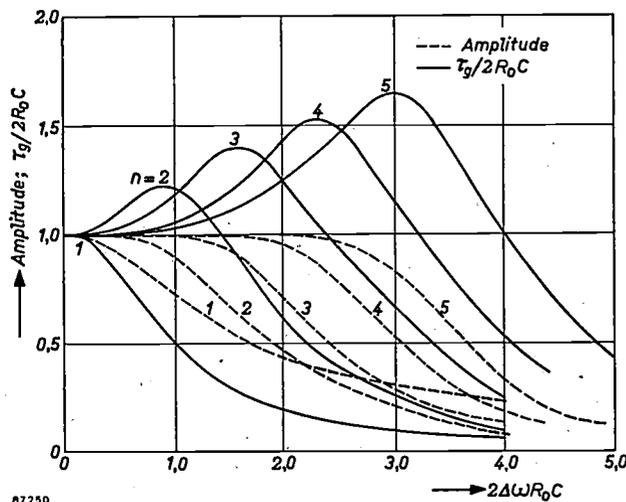


Fig. 12. Amplitude and group delay curves over one half of the band admitted by a wide-band amplifier with  $n$  staggered circuits, for various values of  $n$ . The quantity  $R_0C$  is a measure of the average impedance of the  $n$  circuits. The aim in view was to attain as flat an amplitude curve as possible.

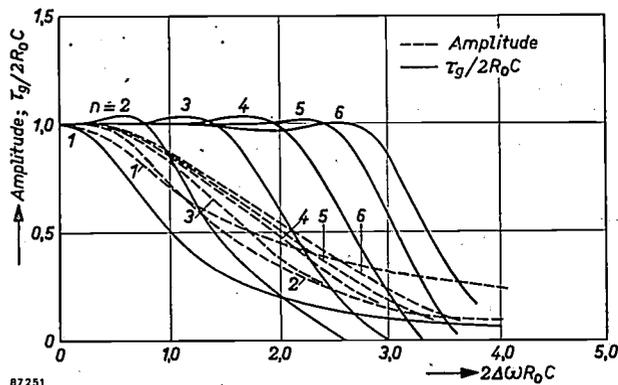


Fig. 13. The same as fig. 12, except that the aim in view was as constant a group delay as possible.

of fig. 14, in which the crosses denote the case with three LC elements and the circles the case with six elements. The full lines drawn in the figure are parabolae, the parameters of which are chosen such that the curves agree as closely as possible with the transferred amplitude curves for low abscissa values. With three LC elements, marked divergence from the parabola begins at an abscissa value of 1.6; with six LC elements it begins at an abscissa value of 2.8. We see from fig. 13 that these are indeed the values at which the group delay curves begin to deviate considerably from flatness.

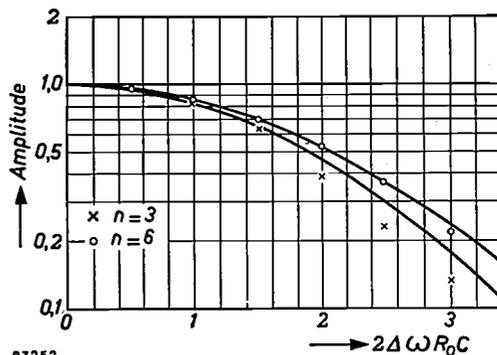


Fig. 14. Crosses: Points of the amplitude characteristic for  $n = 3$  in fig. 12. Circles: the same for  $n = 6$ . The full curves are the parabolae that most nearly fit the crosses and circles at low abscissa values.

### The Gaussian curve as an IF selectivity curve

In order to investigate the suitability of a Gaussian function as amplitude characteristic for an IF amplifier, we have drawn such a function on a logarithmic scale (the parabola  $I_a$ ) in fig. 15a. The parameter of the parabola has been chosen such that the amplification at the frequencies both of the IF picture carrier (38.9 Mc/s) and of the side-band component corresponding to a modulation frequency of 4.5 Mc/s (i.e. 34.4 Mc/s) falls to one half. It is an easy matter (if the modulation depth

is small) to work out the corresponding amplitude curve for the modulation by reading the values of the two side-band components of each modulation frequency from curve  $I_a$ , and adding each pair. The result is curve  $I_b$  in fig. 15b.

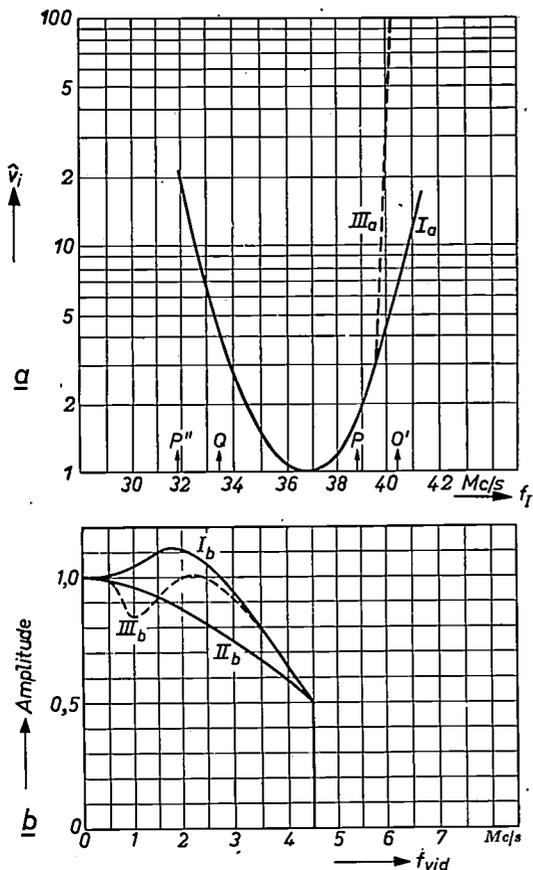


Fig. 15. a)  $I_a$ : selectivity curve of IF amplifier in the form of a parabola (Gaussian function).  $III_a$ : selectivity required for suppression of the adjacent sound carrier  $Q'$ . b) Modulation amplitude characteristics.  $I_b$  corresponds to  $I_a$  in (a),  $III_b$  to  $III_a$ . The desired characteristic is  $II_b$  (cf. curve II in fig. 6a).

Curve  $II_b$  may be taken as an ideal modulation-amplitude curve; it falls a little on approaching the cut-off frequency, to reduce the overshoot factor. It will be seen that there is not a great deal of difference between curves  $I_b$  and  $II_b$ ; if the selectivity it increased (as is required), the disparity becomes even less significant. Disregarding for a moment the possible effect on the delay curve, consider the curve  $III_a$  that does satisfy the selectivity requirement; we see that the corresponding modulation-amplitude curve  $III_b$  approaches the ideal  $II_b$  even more closely. However, we shall not attempt further improvement of the amplitude curve by building on these somewhat hypothetical foundations, but return instead to the subject of phase properties.

The problem now is how to modify the Gaussian selectivity curve in such a way that selectivity

requirements are met while the group delay curve is nevertheless kept flat in the band-pass range. This can be done to quite a close approximation by making use of Bode's relation (eq. (4)), according to which the phase at a given frequency is mainly determined by the shape of the amplitude curve in the immediate vicinity of that frequency. Let us write  $A$ , the attenuation factor in (4), as

$$A = A_1 + A_2,$$

where  $A_1 = a(\omega - \omega_c)^2$ , the amplitude variation is accordance with curve  $I_a$  in fig. 15, and  $A_2$  is the additional attenuation required to bring about the desired suppression of the adjacent sound carrier. The equation for the phase delay then becomes:

$$\frac{\varphi_c}{\omega_c} = \frac{2}{\pi} \int_0^\infty \frac{A_1 - A_c}{\omega^2 - \omega_c^2} d\omega + \frac{2}{\pi} \int_0^\infty \frac{A_2}{\omega^2 - \omega_c^2} d\omega. \quad (6)$$

Now  $A_1$  was a Gaussian function, and consequently the first term on the right-hand side of (6) is again independent of  $\omega_c$ ; thus it is a matter of choosing  $A_2$  such that the second term has only a small effect in the neighbourhood of the picture carrier frequency. One solution of the problem is to make  $A_2$  zero throughout except at those frequencies where additional suppression is necessary, as illustrated in fig. 16. The magnitude of  $A_2$  at these frequencies is determined by the attenuation required; the smaller the band is over which  $A_2$  is not zero, the less will be the effect on phase at the picture carrier frequency, for the value of the second integral in (6) will then be smaller.

Thus, to start with, we may stipulate that suppression must be confined to the narrowest possible band of frequencies, in other words that highly selective trap circuits must be employed.

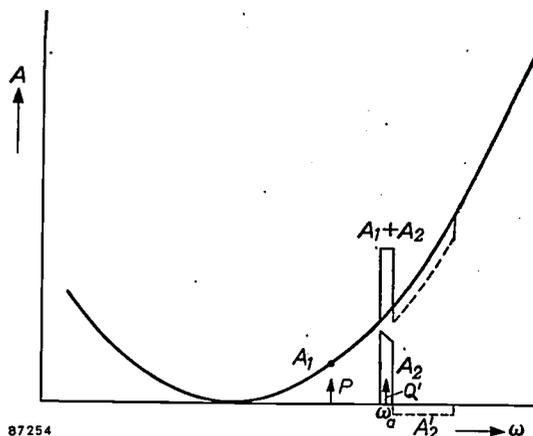


Fig. 16.  $A_1$  attenuation factor of parabolic form, i.e. a Gaussian function.  $A_2$  the additional attenuation necessary for suppressing the adjacent sound carrier  $Q'$ .

A second qualitative requirement can be deduced from this general theory. If we add a negative portion to function  $A_2$ , as shown by a dotted line in fig. 16, the total effect of the second term on the right-hand side of (6) will be further reduced for frequencies around the carrier frequency. Experiments have in fact shown that circuits of which the amplitude curve is re-entrant beyond the suppressed frequencies do possess a group delay curve that is flatter over the pass-band. Clearly it is not desirable to push the selectivity in this frequency region beyond the point of real necessity.

### Design of a phase-linear IF amplifier

The foregoing considerations have been taken as the basis for the design of an IF amplifier (see fig. 17) that satisfactorily fulfils the requirements.

ance to the components forming this filter. But by doing so the series resistance is also increased; to obtain effective suppression of the resonance frequency in spite of this, a bridging resistor  $R$  is employed to provide an additional coupling that is opposite in phase to the remaining coupling via  $T_3$ . In this way the effective  $Q$  of  $T_3$  can be raised to as high a value as desired, at least insofar as the residual coupling at resonance is concerned.

The resistor  $R$  also effectively reduces the damping in the series-resonant circuit  $T_4$ , whose function is to suppress the sound carrier of the desired channel. The degree of suppression must not be excessive, on account of the employment here of the intercarrier sound system; the necessary amount of damping of  $T_4$  is achieved by inserting an extra series resistance.

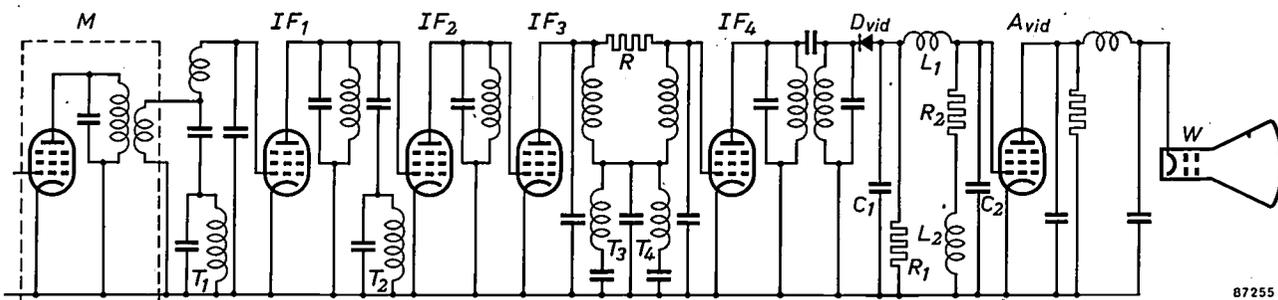


Fig. 17. Essentials of the circuit diagram of a phase-linear receiver.  $M$  mixer stage.  $IF_1$ - $IF_4$  IF stages.  $T_1$  trap for suppressing the adjacent picture carrier (of frequency 31.9 Mc/s,  $P''$  in fig. 4b). The effect on the trap circuit  $T_2$  is to produce a sharp rise in the selectivity curve at 41.4 Mc/s (see curve  $I$  in fig. 18a, extreme right) and thus to improve the selectivity around that frequency. The trap circuit  $T_3$  suppresses the sound carrier of the adjacent channel (40.4 Mc/s,  $Q'$ ) and the trap circuit  $T_4$  suppresses the sound carrier of the channel being received (33.4 Mc/s,  $Q$ ). The additional coupling provided by resistor  $R$  reduces the damping of  $T_3$  and  $T_4$ .

The amplifier consists of five stages in two groups, a group of three stages ( $M$ ,  $IF_1$  and  $IF_2$ ) and a group of two stages ( $IF_3$  and  $IF_4$ ). Each of the groups possesses a flat group delay curve within the pass-band and an amplification that falls by a factor of  $\sqrt{2}$  at the limits of the band.

Apart from the trap circuits, each group contains four tuned circuits. The first group consists of a single stage with a band filter plus two further stages each containing one tuned circuit; these two are staggered. The two stages of the second group are each provided with a band filter.

Suppression of the adjacent and desired sound channels is effected by connecting two series resonant circuits ( $T_3$  and  $T_4$  in fig. 17), tuned to the frequencies concerned, in parallel with the coupling capacitor of the band filter of the last stage but one.

In order to confine the effect of  $T_3$  to a narrow frequency band it is necessary to give high imped-

It will be seen from the circuit diagram that the first three stages are provided with the trap circuits  $T_1$  and  $T_2$ . Their purpose is to give supplementary suppression in a certain range of frequency;  $T_1$  does so at 31.9 Mc/s, i.e. at the frequency of the adjacent picture carrier, and  $T_2$  increases selectivity beyond the adjacent sound carrier.

The resulting selectivity and group delay characteristics are shown in figs. 18a and b respectively as the broken-line curves marked  $I$ ; for comparison, the corresponding curves ( $II$ ) for a receiver of normal design are also given. These latter are the same as those in fig. 7. The difference between the properties of the two amplifiers is evident.

Even more interesting than the curves themselves is, of course, the quality of the picture received. Before discussing this question it is worth mentioning that the delay curves of the remaining receiver sections, i.e. the HF and VF sections, were also

covered in the investigation. The HF section proved, as was indeed to be expected, to have but little influence on the phase properties of the whole.

**Practical results**

*Characteristic curves of a phase-linear and a normal receiver*

Figs 19 and 20 show VF amplitude and VF phase delay characteristics measured on a phase-linear receiver (curves I) and a normal receiver (curves II). During the measurements the receivers were fed with a high frequency signal modulated with a VF voltage of variable frequency. Measurement were taken of the amplitude and group delay characteristics from the modulator to the control electrode of the picture tube; the phase-delay curve has been derived from the group-delay curve by graphical integration. The receivers were tuned exactly to the carrier frequency.

In both diagrams a further curve III is shown; it applies to the normal receiver subsequent to the incorporation in its VF section of the system of coarse phase-compensation mentioned on p. 40. It can be seen that this results in considerable dis-

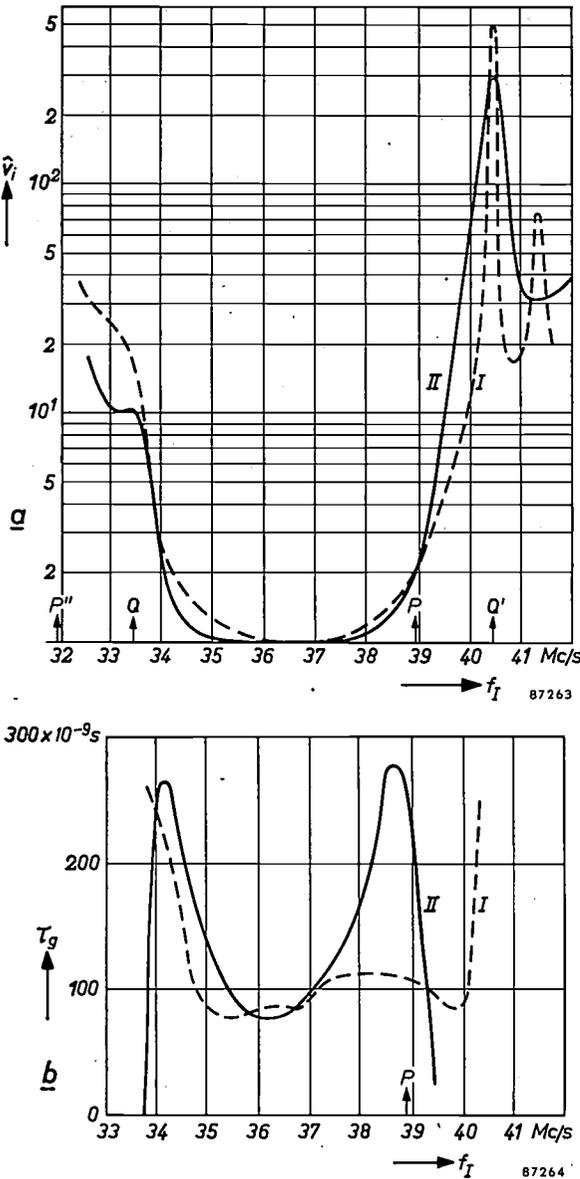


Fig. 18. a) Selectivity curve and (b) group delay characteristic, I for the phase-linear receiver and II for a normal receiver (curves II are taken from fig. 7).

There are few tuned circuits in this section of the receiver; no abrupt changes occur in the shape of the amplitude curve, and there are likewise no great delay variations.

The VF section contains, in the detector circuit, a network whose properties are not theoretically adequately understood<sup>9)</sup>. However, the practical realisation of satisfactory amplitude and delay characteristics for this section offered no particular difficulties.

<sup>9)</sup> Philips Res. Rep. 10, 291-292, 1955 (No. 4).

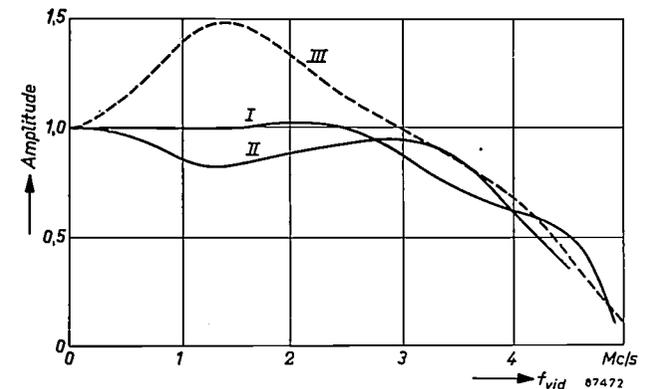


Fig. 19. Amplitude characteristics, I for the phase-linear receiver and II for a normal receiver, III amplitude characteristic for the normal receiver with rough phase compensation in its video frequency section.

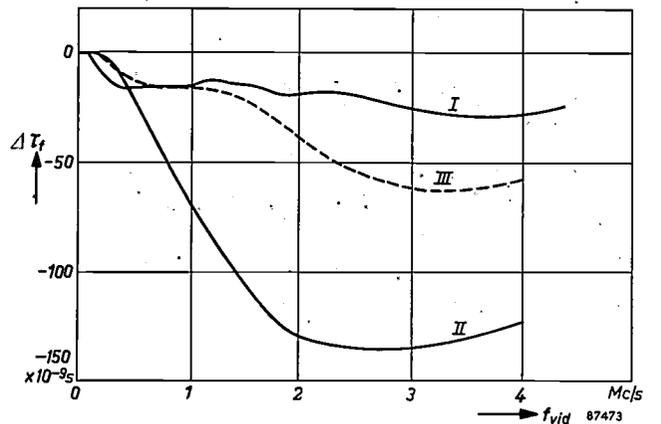


Fig. 20. Phase delay characteristics, I for the phase-linear receiver, II for a normal receiver and III for the normal receiver with rough phase-compensation in its video-frequency section. (ΔT<sub>f</sub> is the difference with respect to the delay-time at low frequencies).

tortion of the amplitude curve while, with regard to phase-linearity, the result is not so good as that of the special receiver.

The characteristics shown in figs 19 and 20 were measured using an HF signal with a double side-band. In fact, the signal of a television transmitter is modulated with an asymmetrical sideband. The rough video frequency curves measured in the latter case differ slightly from those of figs 19 and 20. The difference is not appreciable until frequencies in the region of 1 Mc/s and is more noticeable for the phase-linear receiver than for the conventional receiver. The presence or absence of a second complete sideband has practically no effect on the conventional receiver by reason of the steep slope of the IF amplitude characteristic (see figs 18a and 4), while with a less steep slope the presence of signal components with frequencies near 40 Mc/s has some effect on the video signal after detection.

The rough amplitude curve of a phase-linear receiver measured when a vestigial sideband transmitter is used shows a narrow minimum of 10 to 15% at frequencies near 1 Mc/s. It can be shown that a distortion of this magnitude in the amplitude characteristic is hardly visible in the image<sup>10</sup>); this is also demonstrated by the fact that the pronounced distortion of curve III in fig. 19 is still acceptable. In any case this minimum can be compensated very easily in the video frequency section.

#### Step function response and the effect of modulation depth

Fig. 21 shows the response of the phase-linear receiver to step functions, as recorded by an oscillograph. The high-frequency signal was modulated

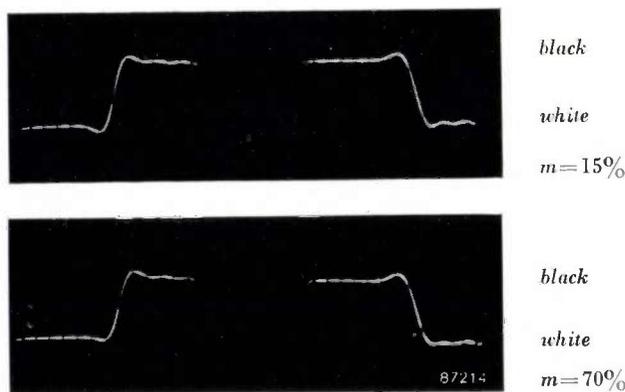


Fig. 21. Step function responses of the phase-linear receiver (oscillograms of the control voltages on the picture tube) for white-to-black transitions (left) and black-to-white transitions (right), the modulation depths  $m$  being 15% in the upper diagram and 70% in the lower one.

with a square-wave voltage to a depth of 15% and 70%; fig. 21 shows the shape of the control voltage on the picture tube for the two cases.

The reason that the modulation depth has an effect is due to the use of an asymmetric side-band system for television. On detection, this produces distortion which is negligible when the modulation depth is small but becomes noticeable at large modulation depth. The kind of distortion produced in the case of a wave modulated with a square-wave voltage is illustrated in fig. 22; it may be seen from

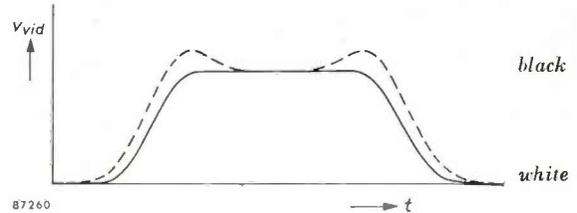


Fig. 22. As a consequence of the distortion arising from the use of an asymmetrical side-band system, the full curve of the step function response takes on the form of the dotted line. The distortion grows with increasing modulation depth and, in the phase-linear receiver, the effect it has can be useful (see fig. 21).

the figure that this effect in itself gives rise to smears with sharp transitions from *black to white* and thus aggravates the effect of phase distortion. The effect of single side-band distortion on a transition from *white to black* is precisely the opposite (fig. 9) and it more or less compensates for the phase distortion.

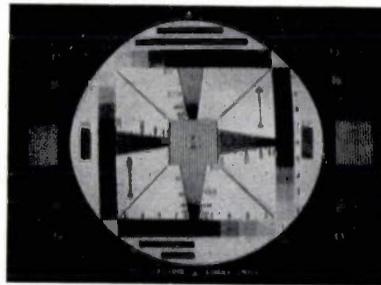
If the oscillograms for 15% modulation depth in fig. 21 are more closely examined, it will be seen that, apart from a slight amount of "ringing", they have a good measure of central symmetry; there is an equal amount of undershoot and overshoot. The central symmetry demonstrates the absence of phase distortion, for the latter deprives the step function of its property of central symmetry (cf. fig. 9). In the oscillogram for 70% modulation depth, it can be seen that the transition from white to black starts sharply; this demonstrates that the undershoot of the curve is compensated here by the distortion due to asymmetrical side-band rectification. On the other hand, at the black end of the transition the overshoot is reinforced; but this has no effect on the received picture, in which "blacker than black" cannot be seen. Something similar occurs for the transition from black to white; at the black end the undershoot is reinforced (which does not matter) and at the white end the overshoot is compensated as a result of single side-band distortion, and this greatly helps to improve the quality of the picture. All in all, the response of the receiver to step functions may be regarded as almost optimum.

<sup>10</sup>) A. van Weel, J. Brit. Instn. Rad. Engrs. 16, 271-282, May 1956.

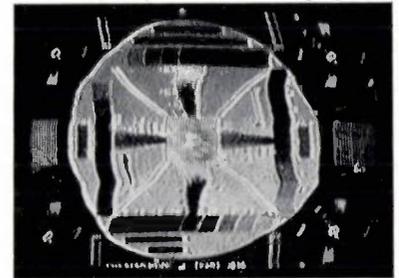
This happy state of affairs is the consequence of two favourable circumstances in combination: phase errors are obviated and, in connection with this, a "Nyquist flank" (the flank of the selectivity curve in the vicinity of the carrier frequency) of moderate steepness is employed (see fig. 18a, curve I). The steeper the Nyquist flank is, the more the asymmetrical side-band system assumes the properties of a pure single side-band system, with the distortion on detection that goes with it. Use of a Nyquist flank that is not so steep gives properties that tend towards those of a double side-band system, at least for modulation components of the lowest picture frequencies, and it is these that are most strongly represented in a step function and that give the greatest modulation depth. In such a case, therefore, asymmetrical side-band distortion is slighter; in this receiver, as we have seen, it is just enough to compensate overshoot and undershoot at the white end of the transition. It is of no importance that, at small modulation depths, undershoot and overshoot phenomena up to about 5% persist at both ends, for 5% represents a fraction of the reduced jump in brightness values that is no longer noticeable.

#### Effect of detuning

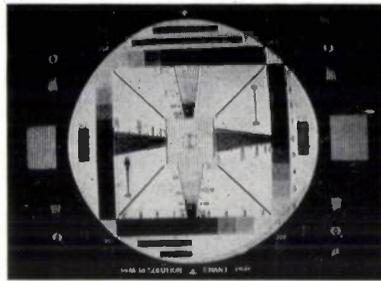
Apart from optimum picture quality, the phase-linear receiver has a very useful property that conventional receivers only possess to a much slighter degree: detuning the receiver is found to have hardly any effect on the picture. With a receiver of the type used for the comparison in figs 18, 19 and 20, the picture quality is very sensitive to the tuning. It is not difficult to see why this should be bearing in mind that, when the conventional receiver is correctly tuned, a satisfactory picture is given only by virtue of the phase compensation in the video section; on detuning, and thereby changing the intermediate frequency of the carrier, the resulting



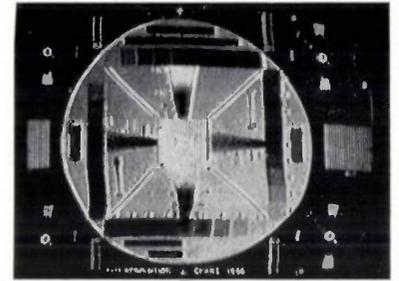
a



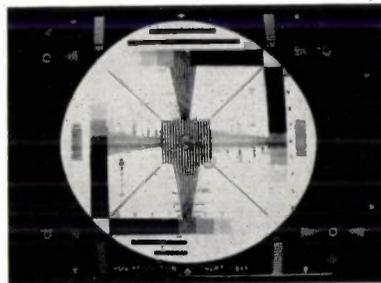
b

Fig. 23.  $\Delta f = +0.55$  Mc/s.

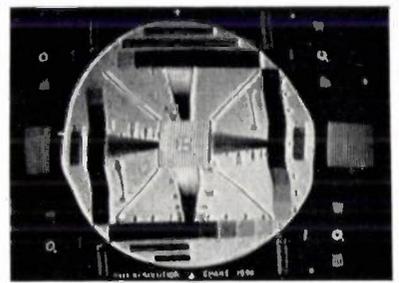
a



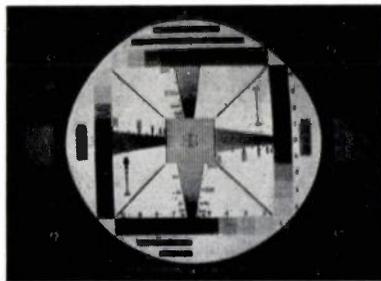
b

Fig. 24.  $\Delta f = +0.50$  Mc/s.

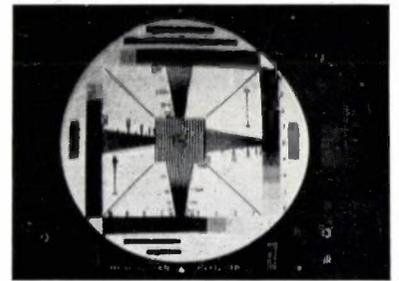
a



b

Fig. 25.  $\Delta f = +0.25$  Mc/s.

a

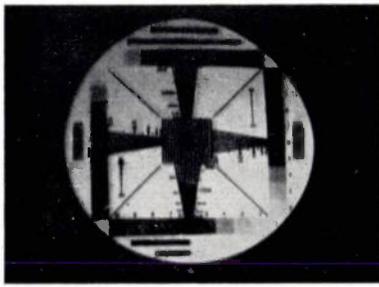


b

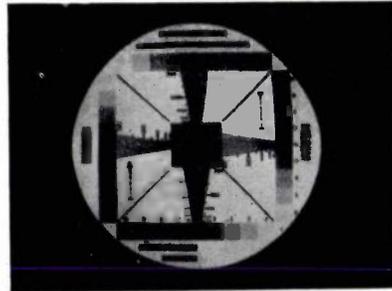
Fig. 26.  $\Delta f = 0$ .

Figs 23-30. Photographs of screen images given by (a) the phase-linear receiver, and (b) a normal receiver having phase compensation in its video-frequency section.  $\Delta f$ : detuning of the transmitter (when  $\Delta f$  is negative, the picture carrier is shifted towards the centre of the pass-band). The generally poor definition of the images is due to their reduced size and to limitations in the printing process; with good tuning (as in fig. 26), it was possible to distinguish more than 400 lines on the screen with the naked eye.

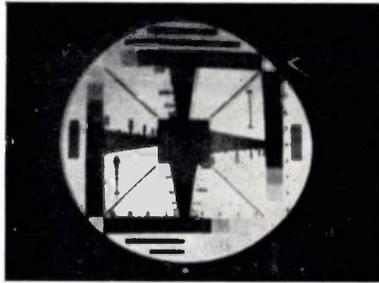
IF phase errors will bring about quite different phase errors in the modulation (see fig. 18b), which are no longer balanced by the phase pre-distortion of the transmitter. This may be seen in figs 23 to 30. These are photographs of the picture screens of both receivers; the tuning of the receivers was left unchanged, the transmitter frequency being altered. The a series of photographs refers to the



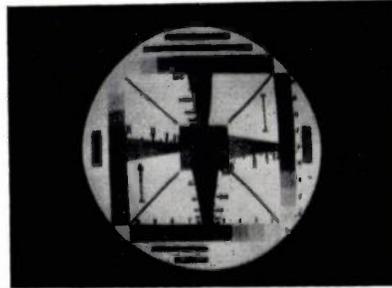
a



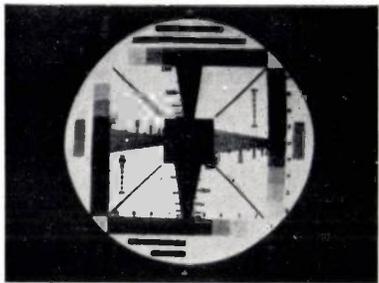
b

Fig. 27.  $\Delta f = -0.25$  Mc/s.

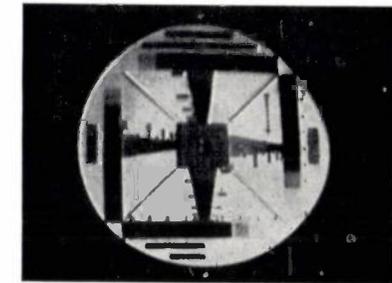
a



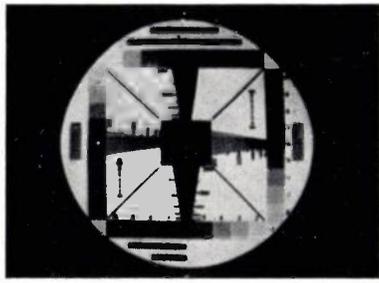
b

Fig. 28.  $\Delta f = -0.50$  Mc/s.

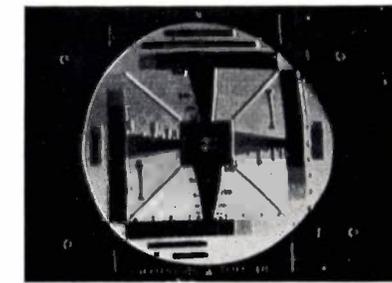
a



b

Fig. 29.  $\Delta f = -1.25$  Mc/s.

a



b

Fig. 30.  $\Delta f = -2.25$  Mc/s.

(The photos of figs. 23-30 were taken during reception of a *double-side-band* transmitter. With reception of a *single-side-band* transmitter, the images given by positive detuning would be the same. For negative detuning — i.e. towards the middle of the pass-band — the effect would be different from that illustrated above, one reason being that the higher picture frequencies would be attenuated by half; this would naturally be the case for both receivers.)

phase-linear receiver and the *b* series to the ordinary receiver.

#### *Phase-linear reception compared with reception using phase compensation*

To conclude this article we shall once again weigh up the advantages and drawbacks of the two practical solutions to the phase problem.

We shall leave the question of phase compensation in the video section out of consideration, since the coarse compensation system described on page 40 cannot provide optimum picture quality, while ideal phase compensation having no effect on the amplitude characteristic cannot be employed in ordinary receivers for economic reasons. We shall confine ourselves, therefore, in making this comparison, to phase compensation in the video-frequency section of the transmitter, and the employment of phase-linear receivers.

One of the most important questions is whether an inherently phase-linear receiver is dearer to produce than a similar receiver of conventional type. This has proved *not* to be the case. The phase-linear receiver described above possesses the same number of stages and tuned circuits in its IF section as the receiver chosen for comparison, and the overall amplification given by its IF amplifier is likewise the same. The alignment of the IF circuits in the phase-linear receiver will be less critical than that in a normal receiver, in which large errors are compensated by intentionally introducing opposite errors, because, in the latter system, small relative variations can produce big absolute errors.

Further striking advantages offered by the phase-linear receiver, as against the system of compensation at the transmitting end, are as follows.

1) Picture quality is much less sensitive to variations in tuning.

2) Optimum picture quality is obtained as a consequence of the reduced effect of vestigial side-band detection.

3) The unfavourable influence of a parasitic phase modulation at the transmitter is less noticeable the flatter the I.F. amplitude curve at frequencies near the carrier frequency. For this reason the phase-linear receiver is less

sensitive to such parasitic phase modulations.

4) No artificial phase distortion needs to be introduced into simple receivers without traps, intended for local reception, in order to compensate pre-distortion at the transmitter, which is pointless in such circumstances.

5) If phase pre-distortion can be dispensed with at the transmitter, the modulation is less difficult: the pre-distortion (or "pre-emphasis") means sending out "whiter than white" to compensate for the black smears in the received picture, and therefore it necessitates a greater maximum modulation depth in the transmitter. This increased modulation depth increases the distortion which accompanies detection of an asymmetrical side-band signal.

6) Both in the production of television receivers and in servicing workshops, picture quality is usually checked with the help of a standard signal generator, the signal from which is double side-band modulated with a testing signal. If the receiver's phase errors are compensated at the transmitter, corresponding phase predistortion will have to be incorporated in the testing generator. This complication does not affect the production of phase-linear receivers, and a much simpler generator can therefore be used<sup>10</sup>).

One disadvantage of the phase-linear receiver discussed here is that the suppressed frequency band is narrow (fig. 18a). Consequently, if interference from an adjacent transmitter is present, fairly high demands have to be made as to the stability of the local oscillator. That is to say, the frequency of the interfering signal in the IF amplifier must never be allowed to deviate from that to which the trap circuit is tuned.

Hence the aim of continuing development work is to widen the suppressed band. It may be mentioned that, at the time of writing, success has been obtained in extending the suppressed band to about 300 kc/s, with attenuation to 1%. Such a width allows for acceptable tolerances as regards the stability of the local oscillator. The circuit necessary for this increased width of the suppressed band is not complicated and does not involve higher production costs.

Another objection that might be made to the phase-linear receiver described here concerns the falling off of the amplitude characteristic at high modulation frequencies. As mentioned above, a compromise must be sought here between greater steepness of the step function response and a higher degree of overshoot. We settled for a falling off of the amplitude characteristic by one half at the edge of the pass-band. However, there are receiver

designers who want the amplitude curve to run flat right up to the cut-off frequency and who accept the correspondingly high overshoot factor (9%) for the sake of obtaining a flank of somewhat steeper slope. It is quite feasible to put this principle into practice; but then, of course, the group delay curve at high modulation frequencies will start to climb earlier and more rapidly. This however has little effect on the phase delay of the modulation, as already stated. The delay curve shown in fig. 8b was measured on an IF amplifier that possesses a flat amplitude characteristic up to 34.4 Mc/s; thus its video-frequency amplitude characteristic is flat up to 4.5 Mc/s. As may be seen from fig. 8b, considerable deviations in group delay have but little effect on the phase delay. An amplitude characteristic of this kind can therefore be realized fairly simply. Difficulties as to phase-linearity are mainly to be expected in the vicinity of the carrier frequency.

In general, the obviation of errors is to be preferred to their compensation by the artificial introduction of opposite ones, unless this offers definite advantages. In the present instance we know of no advantages of this latter practice, but there are several disadvantages as have been outlined above. We believe therefore that the transmitter should radiate a signal that contains as little amplitude and phase distortion as possible, and receivers should likewise be designed such that a minimum of amplitude and phase distortion arises in them. This point of view seems to us to be not only logical but also quite capable of being put into practice; moreover, taking the long view, it does not tie one down to the present-day stage of development. An additional advantage is that it does away with the need to supplement the Gerber Standard with an internationally laid down system of phase compensation. There is no mention of phase properties in the Gerber Standard as at present formulated. This may be interpreted, with regard to the future, as a stipulation that both transmitter and receiver should be phase-linear, in the same way that both should exhibit a flat amplitude characteristic.

---

**Summary.** The growing number of television transmitters has made it necessary to increase the selectivity of receivers especially in the IF section. This has entailed considerable deviations from the linear relationship between phase and frequency necessary to provide a picture of good quality. Up to now a solution of this problem has been sought in the compensation of phase errors, either in the video section of the receiver (where, to give optimum results, it would be very expensive) or in the video section of the transmitter. Phase compensation at the transmitter answers the purpose to only a limited extent,

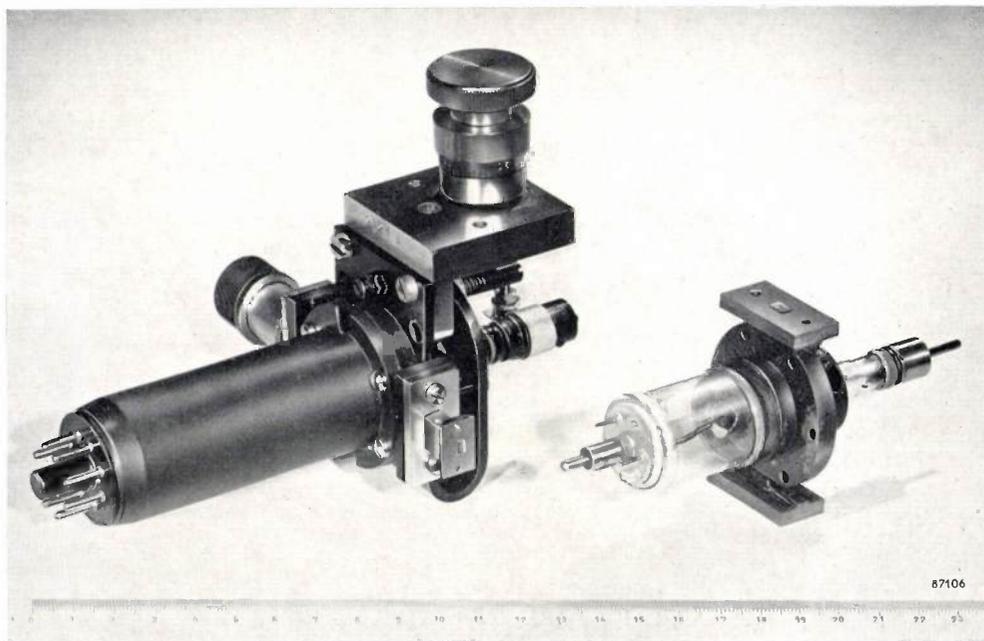
for the various types of receiver differ widely in their phase properties.

Instead of compensation by the introduction of opposite errors, the writer argues for a system in which phase errors can be largely avoided. An IF amplifier was aimed at which would not only possess an optimum amplitude characteristic within the pass-band and satisfy demands as to selectivity outside it, but also be very nearly phase-linear within the band. The aim was to realize such an amplifier without an increase in the cost of manufacture. This aim has been achieved, making

use of Bode's theorem with a 5-stage amplifier, the amplitude curve of which falls to 50% of maximum amplification at the limits of the passband; the requisite selectivity at frequencies capable of causing interference is provided by four trap circuits. In this way optimum picture quality has been obtained for the given bandwidth. An additional favourable property of the phase-linear receiver lies in the fact that its picture quality is much less sensitive to detuning than is the case in an ordinary receiver; this is demonstrated by photographs of screen images.

### A REFLEX KLYSTRON FOR 4-mm WAVES \*)

621.373.423.029.65



In the wake of centrimetre waves, which in the past twenty years have been widely applied, particularly in the field of radar, it is now the range of millimetre waves that is commanding increasing interest. This interest is partly attributable to the wish to improve the "resolving power" of short-range radar installations, partly also to the desire to explore this field with a view to new possible applications.

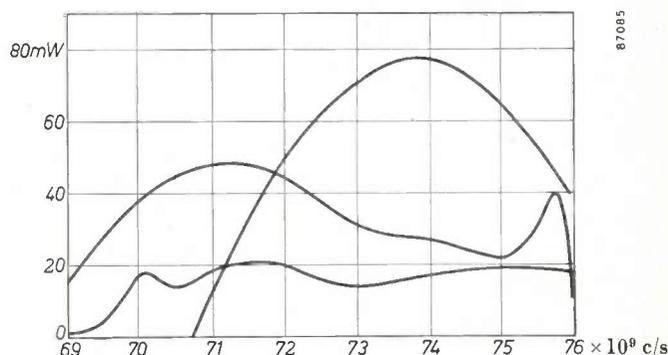
The shortest wavelength employed in practice so far is 8 mm. This restriction is mainly imposed by the limitations of the available oscillator tubes.

A considerable step forward in the attempt to open up the range of millimetric waves for practical application has been made at the Eindhoven Research Laboratories by the development last year of a reflex klystron for 4-mm waves, giving

a continuous output of some tens of milliwatts.

The photograph (scale in centimetres) shows on the right the klystron itself and on the left an assembly fitted with a tuning device, matching plunger, output circuit and base.

The principal difficulties in the construction of tubes for these extremely short waves are due to the necessarily very small size of all component parts.



\*) This note first appeared in the Dutch edition of this review: Philips Technisch Tijdschrift 18, 55-56, February 1956.

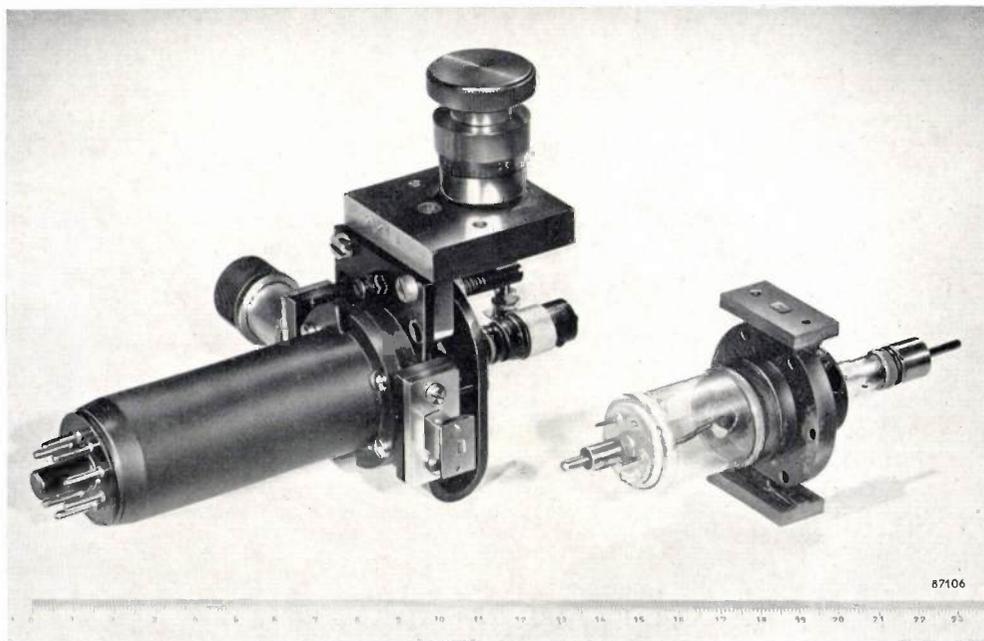
for the various types of receiver differ widely in their phase properties.

Instead of compensation by the introduction of opposite errors, the writer argues for a system in which phase errors can be largely avoided. An IF amplifier was aimed at which would not only possess an optimum amplitude characteristic within the pass-band and satisfy demands as to selectivity outside it, but also be very nearly phase-linear within the band. The aim was to realize such an amplifier without an increase in the cost of manufacture. This aim has been achieved, making

use of Bode's theorem with a 5-stage amplifier, the amplitude curve of which falls to 50% of maximum amplification at the limits of the passband; the requisite selectivity at frequencies capable of causing interference is provided by four trap circuits. In this way optimum picture quality has been obtained for the given bandwidth. An additional favourable property of the phase-linear receiver lies in the fact that its picture quality is much less sensitive to detuning than is the case in an ordinary receiver; this is demonstrated by photographs of screen images.

### A REFLEX KLYSTRON FOR 4-mm WAVES \*)

621.373.423.029.65



In the wake of centrimetre waves, which in the past twenty years have been widely applied, particularly in the field of radar, it is now the range of millimetre waves that is commanding increasing interest. This interest is partly attributable to the wish to improve the "resolving power" of short-range radar installations, partly also to the desire to explore this field with a view to new possible applications.

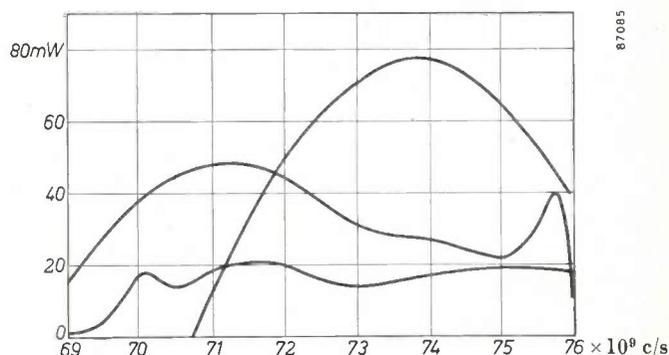
The shortest wavelength employed in practice so far is 8 mm. This restriction is mainly imposed by the limitations of the available oscillator tubes.

A considerable step forward in the attempt to open up the range of millimetric waves for practical application has been made at the Eindhoven Research Laboratories by the development last year of a reflex klystron for 4-mm waves, giving

a continuous output of some tens of milliwatts.

The photograph (scale in centimetres) shows on the right the klystron itself and on the left an assembly fitted with a tuning device, matching plunger, output circuit and base.

The principal difficulties in the construction of tubes for these extremely short waves are due to the necessarily very small size of all component parts.



\*) This note first appeared in the Dutch edition of this review: Philips Technisch Tijdschrift 18, 55-56, February 1956.

The separation between reflector and resonator gap, for example, is only 0.2 mm in the 4-mm tube, across which a voltage of 2.5-3 kV has to be applied. The current-density required of the cathode, which has an emitting surface only about 1 mm in diameter, is roughly 2 A/cm<sup>2</sup>. The construction of such a tube, with a reasonable working life, was made feasible only by the use of the L-cathode. A

more comprehensive description of the tube will shortly appear in this Review.

The graph shows the output power plotted as a function of the frequency for three specimens of the tube. The disparity in the results is admittedly still fairly large, but the power outputs attained at bandwidths of 7-10%, are ample for a number of applications.

B. B. van IPEREN.

## THE STUDY OF THERMAL CONDUCTIVITY PROBLEMS BY MEANS OF THE ELECTROLYTIC TANK

by F. REINIGER \*).

536.212:621.386.1.032.22:621.317.329

*The electrolytic tank, a well-known aid in the study of electrostatic fields, may also be successfully applied for the study of steady state thermal conductivity problems. This article describes an investigation, using this technique, into the thermal behaviour of an oil-cooled copper X-ray anode with a small thin plate (pastille) of tungsten embedded in it.*

It is well-known that the form of electric equipotential surfaces between an arbitrary electrode system can be determined by means of the electrolytic tank <sup>1)</sup> <sup>2)</sup> <sup>3)</sup>. This has shown itself to be of great practical use for electrode systems which are of such an involved nature that direct calculation of the field is impracticable. Fig. 1 shows a possible arrangement in diagrammatic form. A sinusoidal

alternating voltage originating from the oscillator *I* is applied to the electrodes  $E_0$  and  $E_1$  of the electrolytic tank and a potentiometer  $P_0P_1$  connected in parallel. The probe electrode *S* is dipped in the electrolyte and the position of the sliding contact on the potentiometer is adjusted so that the amplitude of the alternating voltage on this contact is equal to that on the probe. Keeping the sliding contact fixed, the probe is then moved along the surface of the liquid so that its potential remains the same. The probe then describes an equipotential line which can be recorded on a drawing board set up near the tank by means of a stylus coupled mechanically to the probe. The method best lends itself to the study of two-dimensional problems but it can quite simply be extended to cover three-dimensional ones with a plane of symmetry, the surface of the liquid in this case having to coincide with the plane of symmetry.

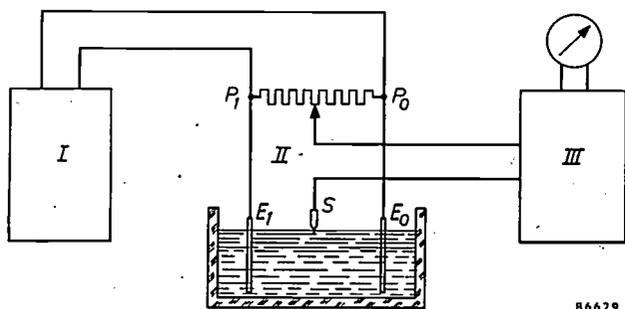


Fig. 1. Diagram of an arrangement for measuring the potential in an electrolytic tank. *I* oscillator, *II* tank and potentiometer, *III* amplifier with meter.

\*) C. H. F.-Müller Aktiengesellschaft, Hamburg. The investigation described in this article partly refers to experiments carried out in 1945 but about which nothing has so far been published.

<sup>1)</sup> G. Hepp, Potential measurements by means of the electrolytic tank, Philips techn. Rev. 4, 223-230, 1939.

<sup>2)</sup> F. Möller, Arch. techn. Messen, V24-1, July 1947.

<sup>3)</sup> N. Warmoltz, Potential distribution of the igniter of a relay valve with mercury cathode, Philips tech. Rev. 8, 346-352, 1946.

The study of electrostatic fields by means of the tank rests on the analogy between the dielectric displacement (vector **D**) in an electrostatic field and the current density (vector **j**) in a conducting medium. For the component of **D** and **j** in an arbitrary direction (*l*) we have the equations:

$$D_l = \epsilon E_l = -\epsilon \partial u / \partial l \dots \dots \dots (1)$$

and

$$j_l = \gamma E_l = -\gamma \partial u / \partial l \dots \dots \dots (2)$$

The separation between reflector and resonator gap, for example, is only 0.2 mm in the 4-mm tube, across which a voltage of 2.5-3 kV has to be applied. The current-density required of the cathode, which has an emitting surface only about 1 mm in diameter, is roughly 2 A/cm<sup>2</sup>. The construction of such a tube, with a reasonable working life, was made feasible only by the use of the L-cathode. A

more comprehensive description of the tube will shortly appear in this Review.

The graph shows the output power plotted as a function of the frequency for three specimens of the tube. The disparity in the results is admittedly still fairly large, but the power outputs attained at bandwidths of 7-10%, are ample for a number of applications.

B. B. van IPEREN.

## THE STUDY OF THERMAL CONDUCTIVITY PROBLEMS BY MEANS OF THE ELECTROLYTIC TANK

by F. REINIGER \*).

536.212:621.386.1.032.22:621.317.329

*The electrolytic tank, a well-known aid in the study of electrostatic fields, may also be successfully applied for the study of steady state thermal conductivity problems. This article describes an investigation, using this technique, into the thermal behaviour of an oil-cooled copper X-ray anode with a small thin plate (pastille) of tungsten embedded in it.*

It is well-known that the form of electric equipotential surfaces between an arbitrary electrode system can be determined by means of the electrolytic tank <sup>1)</sup> <sup>2)</sup> <sup>3)</sup>. This has shown itself to be of great practical use for electrode systems which are of such an involved nature that direct calculation of the field is impracticable. Fig. 1 shows a possible arrangement in diagrammatic form. A sinusoidal

alternating voltage originating from the oscillator *I* is applied to the electrodes  $E_0$  and  $E_1$  of the electrolytic tank and a potentiometer  $P_0P_1$  connected in parallel. The probe electrode *S* is dipped in the electrolyte and the position of the sliding contact on the potentiometer is adjusted so that the amplitude of the alternating voltage on this contact is equal to that on the probe. Keeping the sliding contact fixed, the probe is then moved along the surface of the liquid so that its potential remains the same. The probe then describes an equipotential line which can be recorded on a drawing board set up near the tank by means of a stylus coupled mechanically to the probe. The method best lends itself to the study of two-dimensional problems but it can quite simply be extended to cover three-dimensional ones with a plane of symmetry, the surface of the liquid in this case having to coincide with the plane of symmetry.

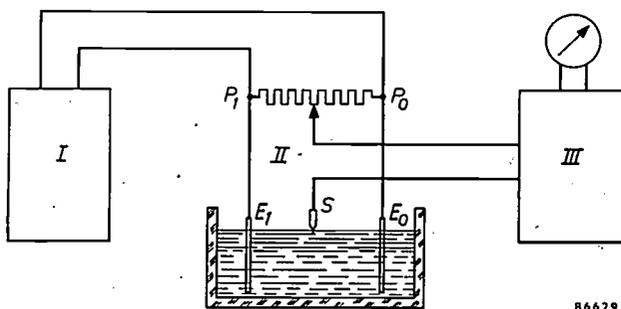


Fig. 1. Diagram of an arrangement for measuring the potential in an electrolytic tank. *I* oscillator, *II* tank and potentiometer, *III* amplifier with meter.

\*) C. H. F.-Müller Aktiengesellschaft, Hamburg. The investigation described in this article partly refers to experiments carried out in 1945 but about which nothing has so far been published.

<sup>1)</sup> G. Hepp, Potential measurements by means of the electrolytic tank, Philips techn. Rev. 4, 223-230, 1939.

<sup>2)</sup> F. Möller, Arch. techn. Messen, V24-1, July 1947.

<sup>3)</sup> N. Warmoltz, Potential distribution of the igniter of a relay valve with mercury cathode, Philips tech. Rev. 8, 346-352, 1946.

The study of electrostatic fields by means of the tank rests on the analogy between the dielectric displacement (vector **D**) in an electrostatic field and the current density (vector **j**) in a conducting medium. For the component of **D** and **j** in an arbitrary direction (*l*) we have the equations:

$$D_l = \epsilon E_l = -\epsilon \partial u / \partial l \dots \dots \dots (1)$$

and

$$j_l = \gamma E_l = -\gamma \partial u / \partial l \dots \dots \dots (2)$$

Table. Corresponding quantities in the electric field, the thermal field and the electrostatic field

Electrical	Thermal	Electrostatic
Current $I$ [A]	Heat flow $N$ [W]	Charge $Q$ [As]
Current density $j$ [A/m <sup>2</sup> ]	Thermal current density $q$ [W/m <sup>2</sup> ]	Displacement $D$ [As/m <sup>2</sup> ]
Potential $U$ [V]	Temperature $T$ [°C]	Potential $U$ [V]
Field strength $E$ [V/m]	Temp. gradient $\tau$ [°C/m]	Field strength $E$ [V/m]
$U - U_0 = u$ [V]	$T - T_0 = \vartheta$ [°C]	$U - U_0 = u$ [V]
Conductivity $\gamma$ [(A/m <sup>2</sup> )/(V/m)]	Conductivity $\lambda$ [(W/m <sup>2</sup> )/(°C/m)]	Dielectric const. $\epsilon$ [(As/m <sup>2</sup> )/(V/m)]
Specific. res. $\rho$ [(V/m)/(A/m <sup>2</sup> )]	—	—
Conductance $G$ [A/V = $\Omega^{-1}$ ]	Conductance $L$ [W/°C]	Capacitance $C$ [As/V]
Resistance $R$ [V/A = $\Omega$ ]	Resistance $R_w$ [°C/W]	—
—	Heat transfer coefft. $\alpha$ [W/m <sup>2</sup> °C]	—

respectively, where  $\epsilon = \epsilon_0 \epsilon_r$  is the dielectric constant of the medium,  $E$  the field strength,  $u = U - U_0$  the voltage with respect to a point at potential  $U_0$  and  $\gamma$  the conductivity of the medium.

It is perhaps less well known that it is also possible to investigate steady state problems of heat flow by means of the electrolytic tank <sup>4)</sup> on the basis of the analogy between the expression (2) and that for the flow of heat  $q$ . Considering again the component in the direction  $l$ ,

$$ql = \lambda \tau l = -\lambda \partial \vartheta / \partial l \dots (3)$$

In this expression  $\lambda$  is the coefficient of thermal conductivity of the medium,  $\vartheta = T - T_0$  the difference between the local temperature  $T$  and a chosen, fixed temperature  $T_0$  and  $\tau$  the temperature gradient  $= -\partial \vartheta / \partial l$  (in the direction of the most rapid decrease).

To clarify the analogy and the subsequent considerations of the thermal problem investigated, corresponding quantities (and their units) are given in the Table for the electric field and for the thermal field; for the sake of completeness, the corresponding quantities in the electrostatic field are also shown. The correspondence follows from a comparison of the differential equations for the three cases, as given in the appendix to this article.

The analogy was used in the present instance for investigating the thermal behaviour of the anode of an X-ray therapy tube (type 71771/05, 200 kV, 4 kW continuous load; see fig. 8). The anode consists of a block of copper with a small, thin plate of tungsten on the front of it where the energy absorption takes place. The other side of the copper is cooled with oil and the heat thus conducted away.

The object of this experiment was to provide details of the temperature distribution in the anode, in relation to the thickness of the copper block, and more particularly, of the distribution of the temperature over the surface in contact with the cooling liquid. Since tungsten is a poorer conductor of heat than copper, the dimensions of the tungsten plate affect the temperature distribution. The effect of the size of the focus on the temperature at the focus was also investigated.

#### Design of the model

In its simplest form, the electrolytic tank comprises a trough filled with a liquid in which two electrodes maintained at constant potential  $U = U_0$ ,  $U = U_1$  are dipped. Since  $u = U - U_0$  is analogous to  $Q$ , the method can be used wherever heat is supplied or absorbed by constant temperature surfaces.

If the heat-conducting medium in the original consists of various materials with different values of thermal conductivity (see Appendix, equations (12)-(14), (16)-(17) and (21)-(23), the model must have areas corresponding to these materials having different electrical conductivities i.e. different electrolytes. The potential must be the same on both sides of the boundary between two areas, i.e. adjacent electrolytes must be separated from each other by a wall whose electrical conductivity is large perpendicular to the plane of the wall but whose conductivity is zero along the wall. In two-dimensional cases, use is made of a wall of insulating material provided with metal strips <sup>3) 5)</sup> (fig. 2). At the position of the strips the potential on both sides of the wall is the same, while there is no conductivity along the wall since the strips are

<sup>4)</sup> J. Malavard and J. Miroux, Electrical analogies for Heat Transfer Problems, Proc. 4th Int. Congr. Industr. Heating, Paris 1952; see Eng. Digest 13, 417-420, 1952.

<sup>5)</sup> R. Stachowiak, Störungsfreie Trennwände für zusammengesetzte Elektrolytmodelle. E.T.Z. 62, 441-443, 1941.

insulated from each other. In other cases a wall is used which is pierced by a large number of rivets whose heads make contact with the liquids on both sides (fig. 3).

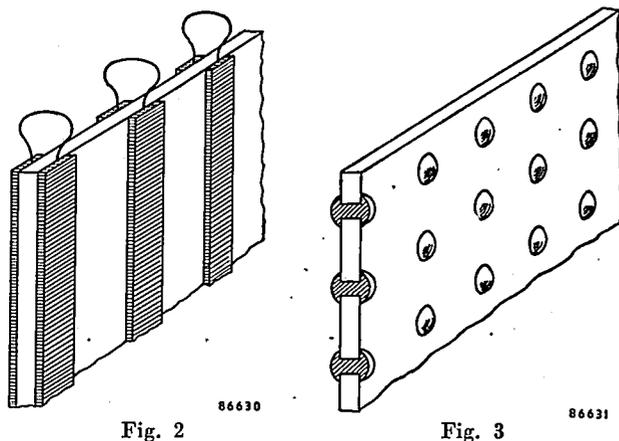


Fig. 2. Construction of a dividing wall with local electrical contact between two electrolytic conductive liquids (two-dimensional).  
Fig. 3. The same as fig. 2 but for a three-dimensional case.

In our particular case another method is used to produce media of different conductivities. Use was made of the principle that a sponge of non-conductive material soaked in an electrolyte is less conductive than the electrolyte liquid itself. The decrease in conductivity resulting from the increase in the current path and the decrease in the conducting cross-section between two points can be measured using the apparatus shown in fig. 4 where the resistance of the electrolyte is determined firstly without, and then with the sponge.

The resistance can be easily increased by a factor of up to 8 by a judicious choice of porosity. There are a number of porous substances suitable for this purpose, some elastic (like rubber) and others non-elastic but all having widely differing porosities (pore volume/total volume). It is of course essential that all pores are interconnected and are distributed isotropically.

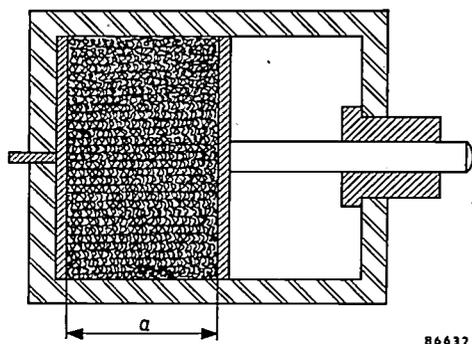


Fig. 4. Device for determining the resistance of an electrolyte and of a sponge-like mass saturated with electrolyte.

It was also possible in the apparatus given in fig. 4 to investigate the manner in which the resistance varies as the sponge is compressed thereby reducing the porosity. The result of this measurement is reproduced in fig. 5. From these results, it can be seen that up to 50 × the original specific resistance can be obtained by compression but care must be taken that the compression does not spoil the isotropy of the sponge. Since it is not easy to work with compressed sponges in a model, this method was not used in the present tests.

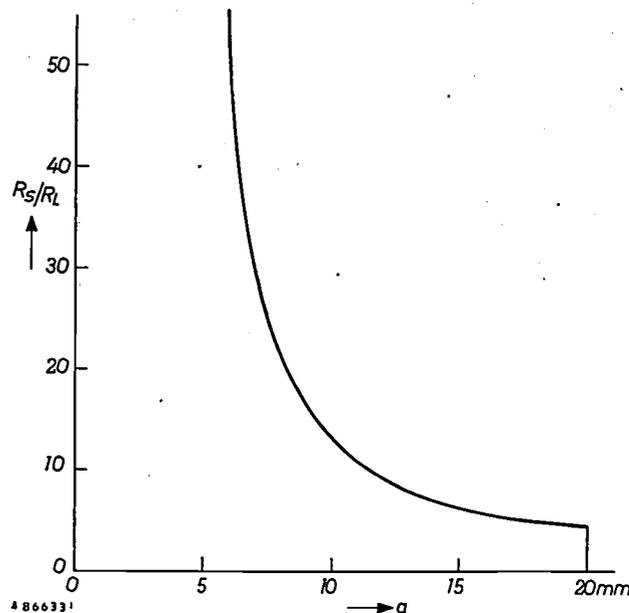


Fig. 5. Variation of the resistance of a sponge saturated with electrolyte with compression. The increase in resistance with respect to the resistance of the electrolyte,  $R_s/R_L$ , is plotted as a function of the length  $a$  in figure 4. In the non-compressed state ( $a = 20$  mm),  $R_s/R_L \approx 4$ .

It is important that the pores, should be completely filled with liquid. Any air-bubbles left behind would give additional resistance to the current. With compressible masses this can be achieved by immersing the sponge completely in the electrolyte and compressing it repeatedly. Non-elastic porous materials can be saturated with liquid in a vacuum and have the advantage of making the model somewhat more rigid.

A condition demanding special attention in the analogy between a thermal and an electric field is the occurrence of temperature discontinuities. In practice, for example, discontinuities occur when a metal wall is being cooled by means of a cooling liquid. There is a layer of liquid which is a bad conductor between the wall to be cooled and the cooling liquid which leads to a temperature drop  $\vartheta_1 - \vartheta_2$ , such that the local heat flow is given by

$$q = a(\vartheta_1 - \vartheta_2) \dots \dots \dots (4)$$

The quantity  $a$  is called the heat transfer coefficient. It is customary to take the temperature of the inflowing cooling liquid for  $\vartheta_2$ . The heat transfer inside the cooling liquid is then included in the constant  $a$ .

A layer of this kind can be simulated in the model by a perforated non-conducting plate (fig. 6) which can be considered as an anisotropic sponge. If the plate contains  $n$  holes per unit area and these holes each occupy an area of  $A = \pi r^2$  ( $r =$  radius of hole) then, if  $\delta$  is the thickness, the current intensity becomes:

$$j = (n\gamma A/\delta)(u_1 - u_2).$$

The quantity  $a$  corresponds to  $n\gamma A/\delta$  and consequently we must ensure that:  $a/\lambda = (n\gamma A/\delta)/\gamma = nA/\delta$ .

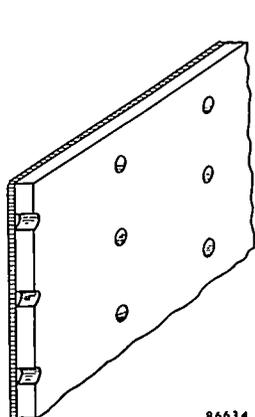


Fig. 6

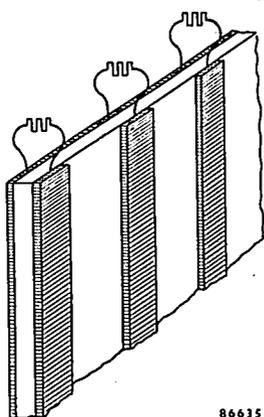


Fig. 7

Fig. 6. Method for obtaining a potential step between the liquid and an electrode by means of a perforated insulating plate covering the electrode (three dimensional). The potential step is the electrical analogy of the temperature drop due to heat transfer across an interface.

Fig. 7. Method for obtaining a potential step between the liquid and an electrode by means of a number of resistances. The arrangement shown is for a two-dimensional case.

Another method of achieving the same thing is to provide the plate with a large number of small metal strips or buttons mounted on an insulating wall (in a two-dimensional case we can again resort to strips, see fig. 7). each connected via external resistors to an electrode on the opposite side of the wall. Although this arrangement is rather complicated with large models, since the resistors must have values adapted to whatever electrolyte is being used, it has the advantage of enabling the value of  $a$  to be adjusted as required and also of allowing it to have a different value at a particular place. (The same can be accomplished with the drilled plate by varying the diameter of the holes.)

As  $a$  depends upon the kind of cooling liquid and upon the state of flow, which is also determined

by the geometry of the cooling space, the value of  $a$  must be ascertained in advance for complicated cooling systems by separate experiments under conditions which simulate as closely as possible those in the original.

The way in which the model is used depends upon the known data e.g. the boundary conditions. If in the original the temperatures at the boundaries be given, then the potentials are adjusted accordingly. For example, let there be two temperatures  $T_1$  and  $T_0$  ( $T_1 > T_0$ ) and let  $U_1$  and  $U_0$  be the corresponding potentials. By ascertaining the equipotential line (in the plane in which the probe moves) information is obtained enabling the temperature at each point to be known. We have:

$$\frac{T - T_0}{T_1 - T_0} = \frac{U - U_0}{U_1 - U_0} \dots \dots \dots (5)$$

or

$$\frac{\vartheta}{\vartheta_1} = \frac{u}{u_1} \dots \dots \dots (6)$$

In other cases (such as the above-mentioned X-ray anode, to which we shall presently refer again in more detail) the total heat flow is known and what is required is the difference in temperature between two points (viz. the focus and the cooling liquid). For this case, it is useful to introduce the conception of thermal resistance (temperature difference/heat flow). In the model, an electrical resistance corresponds to this. Since at a particular point  $x, y, z$  in the original, we have for the  $x$  component of the heat flow:

$$\Delta N_x = q_x \Delta y \Delta z = -\lambda \Delta y \Delta z \cdot \Delta \vartheta / \Delta x, \dots (7)$$

and at the corresponding point  $x', y', z'$  in the model:

$$\Delta I_{x'} = j_{x'} \Delta y' \Delta z' = -\gamma \Delta y' \Delta z' \cdot \Delta u / \Delta x', \dots (8)$$

we obtain

$$\frac{\Delta \vartheta}{\Delta N_x} = -\frac{\Delta x}{\lambda \Delta y \Delta z} \quad \text{and} \quad \frac{\Delta u}{\Delta I_{x'}} = -\frac{\Delta x'}{\gamma \Delta y' \Delta z'}$$

and therefore

$$\frac{R_w}{R} = \frac{\text{thermal resistance}}{\text{electrical resistance}} = \frac{\gamma \Delta x}{\lambda \Delta x'} \cdot \frac{\Delta y' \Delta z'}{\Delta y \Delta z} = m \frac{\gamma}{\lambda},$$

where  $m$  represents the scale ratio:

$$m = \frac{\Delta x'}{\Delta x} = \frac{\Delta y'}{\Delta y} = \frac{\Delta z'}{\Delta z}.$$

### Application to the X-ray Anode

We will now show how the preceding considerations are applied in practice in a particular case. The anode of the X-ray tube mentioned above which

is illustrated in *fig. 8* was the object to be investigated. The electrons strike the focus in the centre of the small tungsten plate  $W$ <sup>6)</sup> and the heat produced is conducted via the tungsten and the copper block Cu to be finally absorbed by the cooling liquid  $K$ .

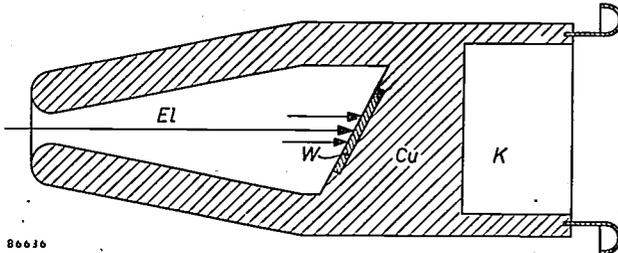


Fig. 8. Cross-section of the anode of an X-ray therapy tube with an oil-cooled anode. Anode potential 200 kV; continuous power rating 4 kW.  $El$  electron stream,  $W$  tungsten plate,  $Cu$  copper body of anode,  $K$  = cooling liquid flowing in turbulent motion.

The object of the investigation is to ascertain the effect on the temperature distribution of the design: in particular the effect of the thickness  $A$  of the anode (distance from centre of focus to cooling surface) and of the temperature at the focus with a given power supply.

On account of the symmetry with respect to the plane of the drawing in *fig. 8*, it is sufficient to consider only half the anode in the model (see *fig. 9a, b*). This is bounded on the outside by a cylindrical surface and since we can disregard the slight loss of heat due to radiation, it is justifiable to represent this surface in the model by an insulating wall. For the latter a sheet of transparent plastic material ("Perspex") bent round to form a half cylinder was used. This half cylinder was enclosed at both ends by a wall of insulating material so that a semi-cylindrical tank was produced and this was filled with electrolyte. The milling on the anode at the side of the tungsten plate was represented by an object (a truncated half cylinder) made of insulating material ("Cellon"). On its oblique face on the cooled side, was fixed a sand-blasted sheet-iron electrode which was to take the place of the focus (the purpose of sand-blasting is to clean the surface and improve the metal-liquid contact).

The model is based on the assumption that the temperature of the focus is uniform at every point but

this is not the case in actual fact. The temperature distribution in the focus depends upon the distribution of the density of the electron stream: At constant current density, the temperature at the centre of the focus will be higher than at the boundary. This detail is not taken into account in the model. It can be said that the potential of the electrode corresponds to the mean focus temperature.

The outer surface of the cooling liquid was represented by a semi-cylindrical electrode, likewise of sand-blasted sheet-iron. On the side where the liquid was, this electrode was completely covered with a perforated, insulating sheet ("Cellon") which represents the heat transfer coefficient. The value of  $nA/\delta$  used in the model was  $\alpha = 0.8 \text{ W/cm}^2 \text{ }^\circ\text{C}$

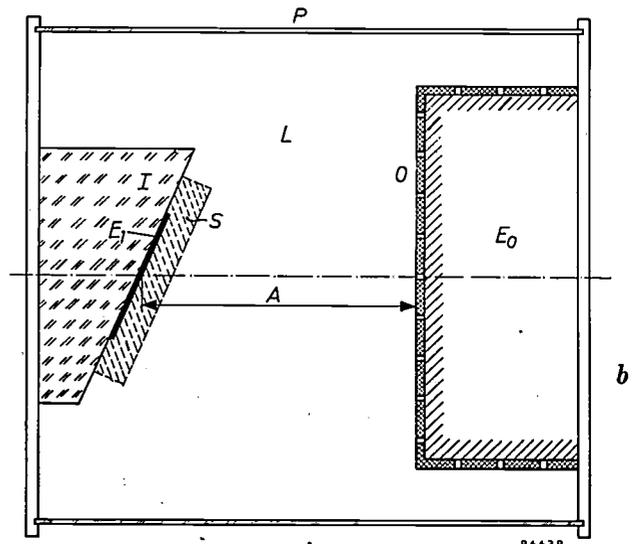
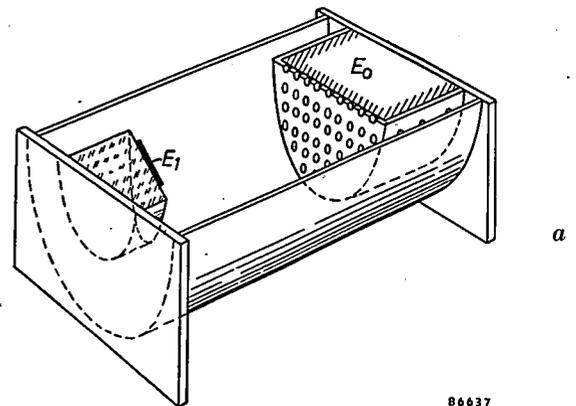


Fig. 9. Perspective sketch (*a*) and cross-section (*b*) of the electrolytic tank model of the X-ray tube anode.  $L$  electrolyte which simulates the heat conducting copper anode.  $P$  semi-cylindrical insulating outside wall of the anode,  $I$  insulator ("Cellon") representing the vacuum in which the electrons move towards the focus,  $E_0$  and  $E_1$  electrodes simulating the cooling liquid and the X-ray focus.  $O$  perforated insulating layer on  $E_0$  representing the heat transfer coefficient  $\alpha$ ,  $S$  = sponge representing the tungsten plate. Three models were prepared, each having a different distance of  $A$ . (In actual fact the iron electrode  $E_0$  and the insulator  $I$  did not need to be made full size but only as thin, conducting and non-conducting limiting walls respectively for the liquid  $L$ ).

<sup>6)</sup> Tungsten is chosen on account of its large X-ray efficiency (high atomic number) and because of its high melting point and low vapour pressure at high temperatures. The relatively poor thermal conductivity of tungsten must be accepted and for this reason the tungsten is in the form of a small, thin plate (pastille) embedded in a substance of good thermal conductivity which may have a lower melting point (electrolytic copper).

(determined from previous experiments concerning the cooling of copper surfaces with oil flowing with turbulent motion).

In the first instance a 3 : 1 scale model was used (scale ratio  $m = 3$ ), the model being in three designs corresponding to anode thickness  $A$  of 35, 25 and 15 mm. Since we are primarily concerned with determining effect of the anode thickness on the temperature distribution at the cooled side, nothing took the place of the tungsten plate in this model. Consequently the model corresponds to an anode consisting solely of copper. The measured equipotential lines for all three cases are drawn in fig. 10. The potential with respect to the cooling liquid is given in percentages of the total potential difference.

thermal conductivity of copper is  $370 \text{ (W/m}^2\text{)/(}^\circ\text{C/m)}$   
We now find

$$R_w = m R \cdot \frac{\gamma}{\lambda} = 3 \times 378 \times \frac{0.0445 \text{ }^\circ\text{C}}{370 \text{ W}} = 0.137 \frac{^\circ\text{C}}{\text{W}}$$

and therefore the thermal conductivity  $L = 1/R_w = 7.3 \text{ W/}^\circ\text{C}$ .

It follows from this that at a load of 4 kW the temperature difference between cooling liquid and focus is:

$$\vartheta_1 = 4000 \times 0.137 \text{ }^\circ\text{C} = 550 \text{ }^\circ\text{C}.$$

At a cooling liquid temperature of  $50 \text{ }^\circ\text{C}$ , temperature  $T_1$  of the focus is accordingly in the region of  $600 \text{ }^\circ\text{C}$ .

The temperature of the copper of the cooling surface can now be easily derived from the per-

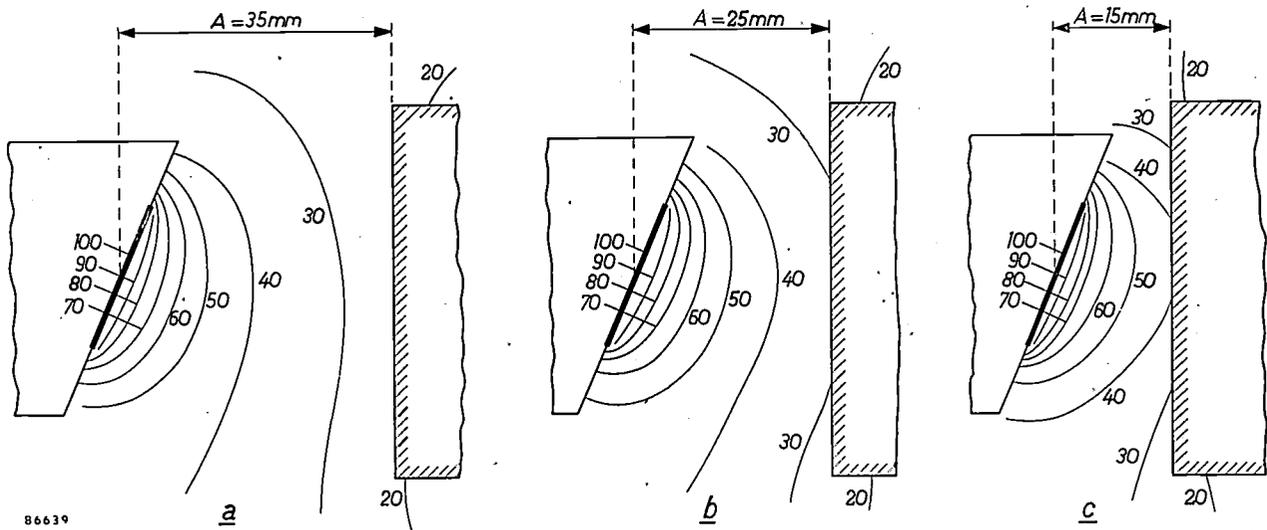


Fig. 10. Equipotential lines measured in a 3 : 1 scale model of the copper anode without the tungsten plate for three thicknesses of the anode: a)  $A = 35 \text{ mm}$ , b)  $A = 25 \text{ mm}$ , c)  $A = 15 \text{ mm}$ .

As can be seen, the rear wall of the anode when 35 mm thick is practically an equipotential surface. In the other cases the temperature varies fairly strongly along the rear wall so that this wall of the anode may in places be hotter than is permissible if decomposition (carbonisation) of the cooling liquid (oil) in the neighbourhood of the copper surface is to be avoided.

The thermal resistance of the anode can be derived from the model tests as follows.

With a thickness of  $A = 35 \text{ mm}$  the electrical resistance  $R$  of the "half" model amounted to 757 ohms. For a complete model it would accordingly be half of this, or 378 ohms. The specific resistance  $\rho = 1/\gamma$  of the electrolyte was measured in an arrangement such as is shown in fig. 4. This came to  $22.5 \text{ } \Omega\text{m}$  and therefore  $\gamma = 0.0445 \text{ } \Omega^{-1}\text{m}^{-1}$ . The

percentage values given in fig. 10. When  $A = 35 \text{ mm}$ , for example, the maximum value of the potential near the cooling surface is about 0.25, so that the temperature of the copper comes to  $50 \text{ }^\circ\text{C} + 0.25 \times 550 \text{ }^\circ\text{C} \approx 185 \text{ }^\circ\text{C}$  (temperature discontinuity approximately  $135 \text{ }^\circ\text{C}$ ).

Later work was done with a 10 : 1 scale model ( $m = 10$ ) and in this model the place of the small tungsten plate was taken by a flat rubber sponge whose pores were saturated with electrolyte. The ratio between the coefficients of thermal conductivity of tungsten and copper at the temperatures which can be anticipated amounts to roughly 3.7:

$$\lambda_{\text{Cu}} = 370 \text{ (W/M}^2\text{)/(}^\circ\text{C/m)},$$

$$\lambda_{\text{W}} = 100 \text{ (W/m}^2\text{)/(}^\circ\text{C/m)}.$$

Care was taken to ensure that the conductivity of

the electrolyte liquid was decreased in the same ratio (approx 1/4) by means of the sponge. In this model, various shapes (rectangular and circular) were used for the focus, i.e. for the electrode  $E_1$ .

In *fig. 11*, the equipotential lines are drawn for an anode 25 mm thick both with and without a tungsten plate. It was possible to obtain the temperature distribution along the normal to the focus from these measurements and the result is reproduced in *fig. 12*. It can be seen from this figure that the temperature curve in the copper with a tungsten plate is practically identical to the one without a plate and furthermore, we can see that the plate makes the temperature of the focus increase strongly. The temperature step (in this instance about 175 °C) between the rear wall of the anode and the cooling liquid can also be clearly seen.

Finally, *fig. 13* gives a résumé of the measurements on various models, with and without a tungsten plate and with different thicknesses of the anode. In these figures, the thermal conductivity

$$L = \frac{\text{power supplied}}{\text{difference in temperature between focus and cooling liquid}}$$

is plotted as a function of the thickness  $A$  (in mm) for various dimensions of the focus (in mm). Obviously, the conductivity is considerably less with the tungsten plate than without it (about half

as much) and the conductivity is greater according as the power supplied is distributed over a greater area (larger focus).

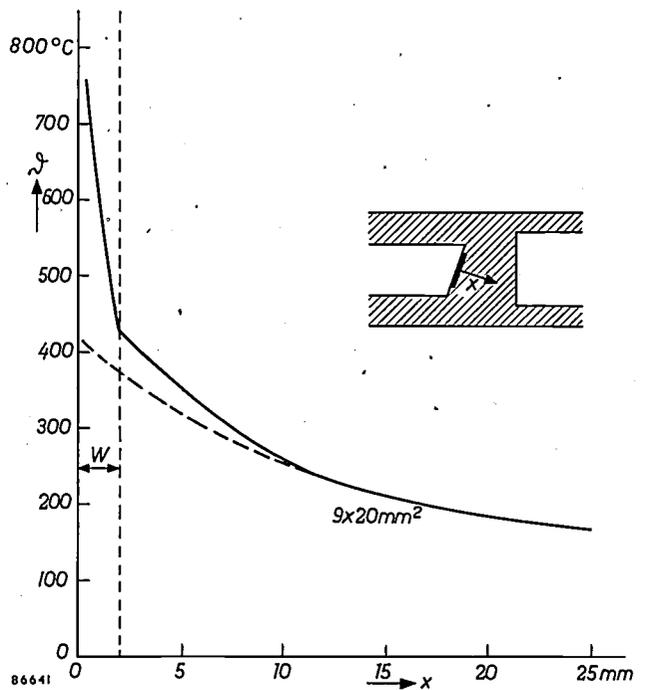


Fig. 12. The temperature curve along the normal in the centre of the focus, obtained from the measurements in *fig. 11*. The temperature difference  $\theta$  with respect to the cooling liquid is plotted. Full line: with tungsten plate; broken line: without plate.

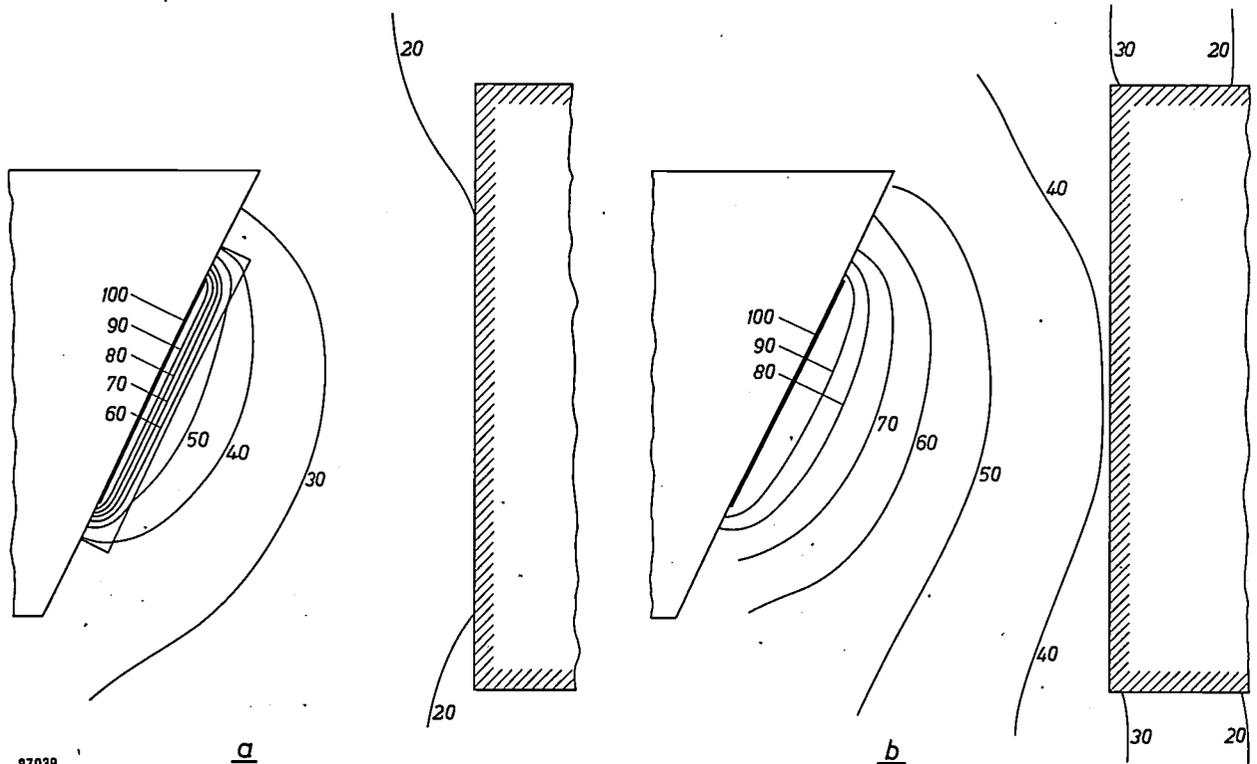


Fig. 11. Equipotential lines measured in a 10 : 1 scale model, a) with and b) without a tungsten plate using an anode of thickness  $A = 25$  mm. The focus dimensions were 9 mm  $\times$  20 mm (the same as in *fig. 10*).

It can be seen from the measurements that the thermal conductivity of the model when the thickness of *A* is reduced, for example from 25 mm to 15 mm, increases but little. On the other hand, as is shown by fig. 10, the maximum temperature at the oil-cooled surface increases still further by about 10% of the temperature difference between focus and cooling liquid. Consequently, by virtue of there being less risk of the cooling liquid carbonizing with a more uniform temperature distribution of the cooled surface, a thicker anode is to be preferred in practice in spite of the lower value of *L* that goes with it.

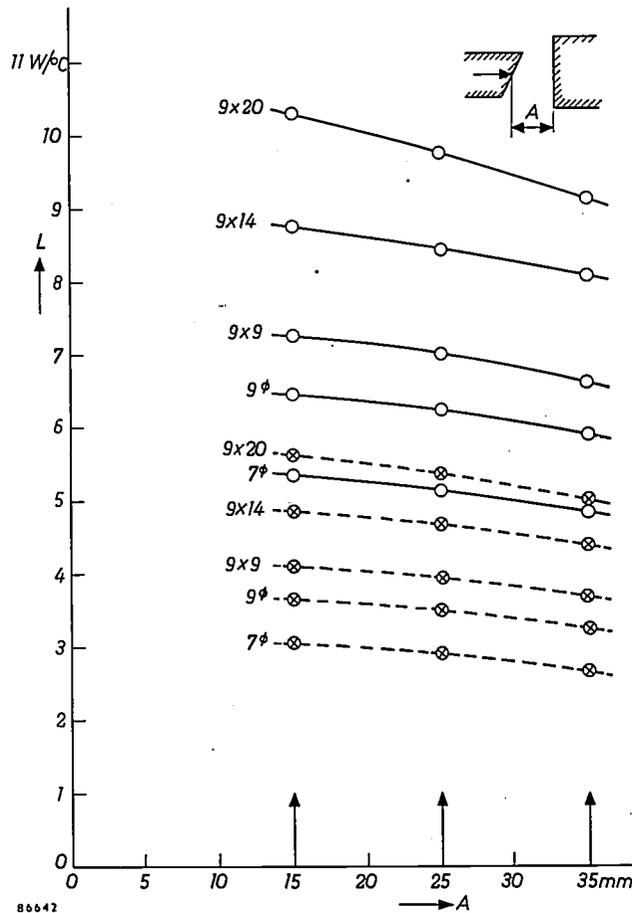


Fig. 13. Thermal conductivity *L* (reciprocal of the thermal resistance) as a function of the thickness *A* for various sizes of the focus (dimensions in mm). Full lines: without tungsten plate; broken lines: with plate.

By means of model tests such as described in this article the behaviour of an X-ray anode can be fairly accurately predicted. In particular, the effect of the area of the focus, the thickness of the anode and of the tungsten plate on the mean temperature of the focus can be found from the tests. The thermal phenomena in the anode are clearly reflected by measurements in the electrolytic tank, these measurements providing a valuable

supplement to theoretical calculations<sup>7)</sup> and excelling them in clarity.

Appendix: Mathematical basis of the analogies used

In the electrostatic field, in the absence of space charge, the following equation applies<sup>8)</sup>

$$\frac{\partial}{\partial x} D_x + \frac{\partial}{\partial y} D_y + \frac{\partial}{\partial z} D_z = 0. \dots (9)$$

If the medium is isotropic and homogeneous ( $\epsilon = \text{constant}$ ), then in accordance with (1) and after division by  $\epsilon$ , we can write instead of (9):

$$\frac{\partial}{\partial x} E_x + \frac{\partial}{\partial y} E_y + \frac{\partial}{\partial z} E_z = 0, \dots (10)$$

or

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0. \dots (11)$$

If the medium is not homogeneous ( $\epsilon$  dependent upon  $x, y$  and  $z$ ), then substituting (10) for (9) is not permissible. If the variation in  $\epsilon$  however is discontinuous so that  $\epsilon$  has the constant values  $\epsilon_1$  and  $\epsilon_2$  at both sides of a specific boundary, then if  $D_n$  represents the component perpendicular to the boundary, we have:

$$(D_n)_1 = (D_n)_2, \dots (12)$$

$$\epsilon_1 (E_n)_1 = \epsilon_2 (E_n)_2, \dots (13)$$

In order that the potential  $U$  has the same value at both sides of the boundary, we have, for the tangential component  $E_t$ :

$$(E_t)_1 = (E_t)_2, \dots (14)$$

For the electrical current intensity  $j$  in a conducting medium, we have, analogous to (9), the equation:

$$\frac{\partial}{\partial x} j_x + \frac{\partial}{\partial y} j_y + \frac{\partial}{\partial z} j_z = 0, \dots (15)$$

which shows that nowhere does a space charge occur as a result of the current. If the medium is homogeneous ( $\gamma = \text{const.}$ ), then by reason of (2) we obtain the same equations (10) and (11) as above. The potential accordingly fulfils the same differential equations as in the electrostatic case.

If  $\gamma$  is a function of  $x, y$  and  $z$ , then substituting (10) for (15) is not permissible. If  $\gamma$  increases from the constant value  $\gamma_1$  to the constant value  $\gamma_2$  at a certain boundary, then for the normal component of  $j$  we have:

$$(j_n)_1 = (j_n)_2, \dots (16)$$

$$\gamma_1 (E_n)_1 = \gamma_2 (E_n)_2, \dots (17)$$

whilst the tangential component again fulfils (14) because the potential has to be equal along both sides of the boundary.

<sup>7)</sup> See, in this connection, W. J. Oosterkamp, The Heat Dissipation in the Anode of an X-ray Tube, Philips Res. Rep. 3, 303-317, 1948. In this publication no account is taken of the temperature discontinuity between the anode metal and the cooling liquid, whereas in the present experiments this is an important variable.

<sup>8)</sup> This can be derived from the equation  $\int D_n dA = Q$ , which holds for an enclosed surface containing the charge  $Q$ . In this,  $D_n$  is the component of  $D$  perpendicular to the surface element  $dA$ . If there is no charge then  $Q = 0$  and if, this being so, we apply the equation to an element with sides  $dx, dy$  and  $dz$ , we obtain equation (9).

From the fact that in both cases we arrive at equation (11), the analogy between the electrostatic field in a homogeneous dielectric and the current field in a homogeneous conductor follows. If the medium in the electrostatic case consists of zones which are individually homogeneous but for which  $\epsilon$  has various values  $\epsilon_1, \epsilon_2$  etc., then we can still work with the model provided we ensure that corresponding zones with  $\gamma$  values  $\gamma_1, \gamma_2$  etc. are present in the tank. It follows from the boundary conditions (13) and (17) that we have to make  $\epsilon_1 = \gamma_1, \epsilon_2 = \gamma_2$  etc.. If we are merely concerned with ratios of potential it is sufficient if  $\epsilon_1 : \gamma_1 = \epsilon_2 : \gamma_2$  etc.

In the case of constant heat flow, we have the equation:

$$\frac{\partial}{\partial x} q_x + \frac{\partial}{\partial y} q_y + \frac{\partial}{\partial z} q_z = 0, \dots (18)$$

which shows that nowhere are there any sources or sinks of heat. If  $\lambda$  is independent of  $x, y$  and  $z$  then on account of (3) we may replace (18) by:

$$\frac{\partial}{\partial x} \tau_x + \frac{\partial}{\partial y} \tau_y + \frac{\partial}{\partial z} \tau_z = 0, \dots (19)$$

or

$$\frac{\partial^2 \theta}{\partial x^2} + \frac{\partial^2 \theta}{\partial y^2} + \frac{\partial^2 \theta}{\partial z^2} = 0. \dots (20)$$

If  $\lambda$  is a function of  $x, y$  and  $z$ , then substitution of (18) by (19) is not permissible. At a boundary between two media where  $\lambda$  increases from the constant value  $\lambda_1$  to the constant value  $\lambda_2$ , we have, for the normal component:

$$(q_n)_1 = (q_n)_2, \dots (21)$$

or

$$\lambda_1 \left( \frac{\partial \theta}{\partial n} \right)_1 = \lambda_2 \left( \frac{\partial \theta}{\partial n} \right)_2. \dots (22)$$

Since the temperature must be equal at both sides of the boundary, we have, for the tangential component:

$$\left( \frac{\partial \theta}{\partial t} \right)_1 = \left( \frac{\partial \theta}{\partial t} \right)_2. \dots (23)$$

From equations (20) and (11) can be seen the analogy between the case of the constant heat flow and the electrical current in a homogeneous conducting medium. The potential  $u$  corresponds to the temperature difference  $\theta$ . When we have various homogeneous media with different  $\lambda$  values, in order that the model should be a replica of the original we must ensure that  $\lambda_1 : \gamma_1 = \lambda_2 : \gamma_2$  etc., as shown by equations (22) and (17).

It must furthermore be stated that  $\gamma$  is substantially independent of  $u$  (Ohm's law) in the electrical case, whereas in general  $\lambda$  does depend slightly on  $T$  and therefore on  $\theta$ . No account is taken of this fact in the model.

---

**Summary.** The analogy existing between the electric field of a potential in an electrolyte and the electrostatic field in a dielectric is mentioned. A similar analogy exists between the potential field in an electrolyte and the temperature field in a conductor under conditions of steady heat flow. Methods of setting up an electrical model (electrolytic tank) for a given thermal conductivity problem are discussed and from this model the required information (temperature distribution in the conductor, thermal resistance) can be obtained. Account must be taken of possible temperature discontinuities e.g. between a cooled metal wall and the cooling liquid. The method is applied to the study of the thermal behaviour of an oil cooled copper X-ray anode carrying an embedded thin plate (pastille) of tungsten. In an appendix, the basis of the analogy of the model is explained by comparing the appropriate differential equations.

## BEHAVIOUR OF THE TRANSISTOR AT HIGH CURRENT DENSITIES

by F. H. STIELTJES and L. J. TUMMERS.

537.311.33:621.375.4

*The current amplification factor of a transistor, which is of great importance for the power amplification obtainable, is found to be dependent upon the current density inasmuch as the current amplification factor falls off if the current density is raised sufficiently. An extension of the theory given in an earlier article, to include those cases where the minority concentrations are not everywhere negligible with respect to the majority concentrations, will account for this effect and indicate how it can be reduced.*

According to the simple theory of the junction transistor as given in an earlier article in this Review <sup>1)</sup>, hereafter referred to as I, the current amplification factor  $\alpha$  of a transistor does not depend upon the applied emitter voltage  $V_e$  (fig. 1), since this voltage does not appear in the approximate for-

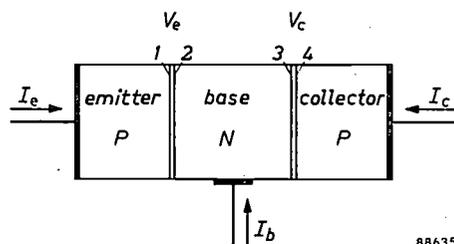


Fig. 1. Schematic representation of a  $P-N-P$  transistor.  $I_e$ ,  $I_c$  and  $I_b$  are respectively the emitter, collector and base currents; 1, 2, 3 and 4 mark the boundary planes of the  $P-N$  and  $N-P$  junctions.  $V_e$  is the voltage between the planes 1 and 2;  $V_c$  is the voltage between the planes 4 and 3 and has a negative value (inverse voltage).

mula derived for  $\alpha$  (I, 13) nor is it found in the more exact expression that can be obtained by the product of formulae (I, 19) and (I, 20). In practice, however, we find that when the current is increased (by raising the emitter voltage) the value of  $\alpha$ , after an initial increase, declines to a value far below its initial value valid for low current densities (see fig. 2, in which  $1-\alpha$  is plotted). The failure of the simple theory is attributable to the fact that one of its premises is no longer tenable when the current density is raised. In fact, as we shall demonstrate, the minority concentration in the base then ceases to be everywhere very small in comparison with the majority concentration at the same spot. This has two important consequences: firstly, the minority current in the base is no longer purely

a diffusion current, for the electric field also begins to have a marked effect; secondly, the expression given in I (I, 1) for the minority concentration in the boundary planes of a  $P-N$  junction is no longer valid. These two facts, which were first pointed out by Webster and by Rittner <sup>2)</sup> will be considered more closely. Subsequently we shall discuss the explanation by these authors of the form of the  $1-\alpha$  curve as shown in fig. 2. From this theoretical explanation we shall deduce that the value of the current density at which the drop of  $\alpha$  becomes troublesome can be substantially raised by the incorporation of a large number of acceptor atoms in the emitter material. In the Research Laboratories at Eindhoven good results have been obtained by the addition of a small amount of gallium to the indium used in the preparation of the emitter of a  $P-N-P$  alloy transistor.

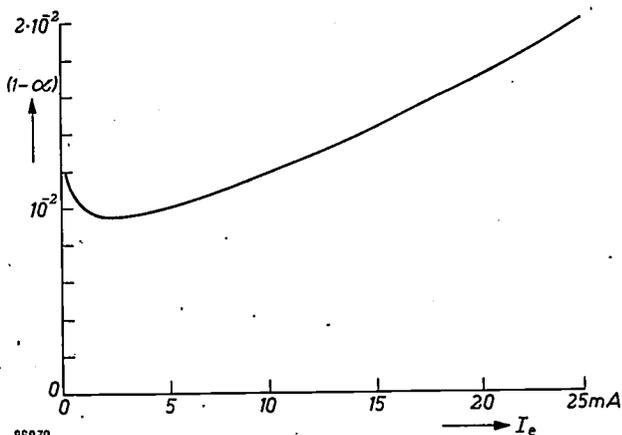


Fig. 2. Only a fraction  $\alpha$  of the total emitter current  $I_e$  reaches the collector.  $1-\alpha$  can, therefore, be considered as a loss. The loss is at a minimum for a relatively low current density, but it climbs continuously as  $I_e$  is further increased. This may be seen from the curve above, which was derived from measurements.

<sup>1)</sup> F. H. Stieltjes and L. J. Tummers, Simple theory of the junction transistor Philips tech. Rev. 17, 233-246, 1955/56 (No. 9). This article may also be referred to for explanations of the usual transistor terminology concerning holes and electrons,  $P$  and  $N$  regions,  $P-N$  junction, majority and minority concentrations, etc.

<sup>2)</sup> W. M. Webster, On the variation of junction-transistor current-amplification factor with emitter current, Proc. I.R.E. 42, 914-920, 1954; E. S. Rittner, Extension of the theory of the junction transistor, Phys. Rev. 94, 1161-1171, 1954.

**Breakdown of the simple theory at high current densities**

*Effect of the field on the minority current at high current densities*

Consider the emitter junction of a P-N-P transistor. The emitter efficiency  $\gamma$ , i.e. the fraction of the emitter current transported by holes, lies close to unity for a good transistor. In the base the current from emitter to collector is therefore almost entirely composed of holes; the electron current is here virtually zero. At every point of the base where the hole concentration (minority concentration) is very small with respect to the electron concentration (majority concentration), that is, at every point where

$$p_N/n_N \ll 1 \dots \dots \dots (1)$$

applies<sup>3)</sup>, the hole current is almost a pure diffusion current. If the hole concentration is so high that (1) is no longer valid, then the electric field also plays a part in determining the hole current.

The proof of these statements, which in this special case (majority current practically zero) is much easier to grasp than the general proof given in I (page 235, small print), is as follows:

As in the general proof, we start from the assumption that hardly any space charge can arise outside a P-N junction (neutrality condition). This implies that the concentration gradients for the two sorts of charge carriers are the same in any cross-section outside a P-N junction (fig. 3). Thus there is not only a diffusion current of holes, but also one of electrons; indeed, the electron diffusion current is about twice as great as the hole diffusion current, for the diffusion constant for electrons is about twice that for holes. Diffusion causes holes and electrons to move in the same direction and hence the electric current densities, designated by  $J_{pD}$  for holes and  $J_{nD}$  for electrons, are of opposite sign so that

$$J_{nD} = -2 J_{pD} \dots \dots \dots (2)$$

If, in spite of this, the total electron current remains (virtually) zero, then there must be an electric field that brings about a field current of electrons (with a density  $J_{nF}$ ) which cancels out the electron diffusion current, i.e.  $J_{nF} = -J_{nD}$ . Replacing  $J_{nD}$  in (2), we can write

$$J_{nF} = 2 J_{pD} \dots \dots \dots (3)$$

If the concentration of electrons is very high compared to that of the holes, the field creates a hole current  $J_{pF}$  which is

<sup>3)</sup> In this article we shall have to take account of the majority concentrations to a far greater extent than was done in I. We must therefore use a somewhat different notation. Hole and electron concentrations will be indicated by  $p$  and  $n$  respectively. The suffix P or N indicates whether a magnitude is related to a P or to an N-region. It is thus immediately apparent whether a minority concentration ( $p_N$  and  $n_P$ ) or a majority concentration ( $p_p$  and  $n_N$ ) is concerned. The corresponding equilibrium concentrations are denoted by an additional suffix 0. The suffixes p and n are used to denote whether a quantity refers to holes or to electrons; e.g.  $J_p$  stands for the hole-current density, and  $D_n$  for the diffusion constant for electrons.

negligible with respect to  $J_{nF}$ , and, in accordance with (3), with respect to  $J_{pD}$  too; if this condition is not satisfied, however, the field current of the holes cannot be disregarded.

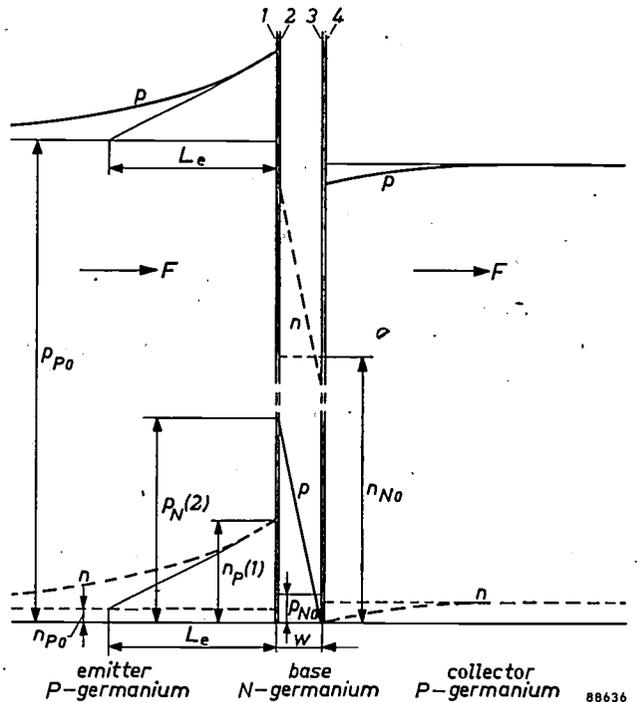


Fig. 3. Hole concentrations (full lines) and electron concentrations (dotted lines) in a P-N-P transistor. 1, 2, 3 and 4 boundary planes of the P-N and the N-P junctions.  $p_{p0}$  and  $n_{p0}$  equilibrium concentrations of holes and electrons in the emitter.  $p_{N0}$  and  $n_{N0}$  the same in the base. The equilibrium concentrations in the collector are not dealt with in this article and are therefore not represented by symbols.  $n_P(1)$  and  $p_N(2)$  minority concentrations in the boundary planes 1 and 2; these are fixed by the voltage applied between 1 and 2. In the boundary planes 3 and 4 the inverse voltage applied between these planes fixes the minority concentrations at zero.  $L_e$  diffusion length in the emitter,  $w$  thickness of base. The electric field  $F$  acts in the direction of the arrows.

The electric neutrality of the base has the consequence that the difference<sup>4)</sup> between the electron concentration  $n_N$  and the hole concentration  $p_N$  is the same at any point and is thus equal to the difference between the equilibrium concentrations  $n_{N0} - p_{N0}$  (fig. 3). If we neglect  $p_{N0}$  in comparison with  $n_{N0}$ , we obtain

$$n_N - p_N = n_{N0} \dots \dots \dots (4)$$

This being so, condition (1) is equivalent to

$$p_N/n_{N0} \ll 1,$$

<sup>4)</sup> One should remember that the nett charge of the mobile charge carriers is balanced by the nett charge of the (fixed) donors and/or acceptors present; electric neutrality does not therefore imply that  $p_N = n_N$ . See eq. (5) in J. C. van Vessem, The theory and construction of germanium diodes, Philips tech. Rev. 16, 213-224, 1954/55.

since, if either of these conditions is satisfied, it follows from (4) that  $n_N \approx n_{N0}$ , so that the other condition is also fulfilled.

If this condition is satisfied in cross-section 2, in other words if

$$p_N(2)/n_{N0} \ll 1, \dots \dots \dots (5)$$

it is evidently satisfied elsewhere in the base, as may be seen from fig. 3. Hence, if we assume that (5) is satisfied, we may conclude that the hole current in the base is almost exclusively a diffusion current. Its current density  $J_p$  is given by:

$$J_p = qD_p \frac{p_N(2)}{w},$$

where  $q$  is the magnitude of the electronic charge,  $D_p$  is the diffusion constant for holes and  $p_N(2)/w$  is the concentration gradient for the holes (fig. 3). Since the electron current is negligible,  $J_p$  is equal to the total current density  $J$ . Hence it follows from (5) that

$$J \ll qD_p \frac{n_{N0}}{w} \dots \dots \dots (6)$$

If, therefore, the current density becomes so high that (6) is no longer satisfied, then condition (5), from which it was deduced, cannot be fulfilled either, and consequently the minority concentration in the base is not everywhere negligible with respect to the majority concentration.

That this situation can occur in practice becomes evident when we insert in (6) the numerical values applicable to a normal *P-N-P* transistor, chosen as an example. (As we shall make repeated use of this same example in the course of this article we shall give some additional data for later reference, apart from that of immediate concern.)

The equilibrium concentrations in the emitter are, say:

$$p_{P0} \approx 5 \times 10^{18} \text{ cm}^{-3}, n_{P0} \approx 10^8 \text{ cm}^{-3},$$

and in the base

$$n_{N0} \approx 5 \times 10^{14} \text{ cm}^{-3}, p_{N0} \approx 10^{12} \text{ cm}^{-3}.$$

Let the base thickness be

$$w \approx 50 \mu = 5 \times 10^{-3} \text{ cm}.$$

The diffusion constants of holes and electrons are

$$D_p = 44 \text{ cm}^2 \text{ sec}^{-1}, D_n = 88 \text{ cm}^2 \text{ sec}^{-1}.$$

The average lifetime of the holes in the base, determined by recombination in the interior and at the surface of the base, is

$$\tau_b \approx 15 \mu\text{sec}.$$

The area of the emitter junction is  $3 \times 10^{-3} \text{ cm}^2$  ( $0.3 \text{ mm}^2$ ). As  $q = 1.6 \times 10^{-19}$  coulomb, we find from (6) that for this particular transistor the influence of the field on the hole current is negligible provided that the current density

$$J \ll 0.7 \text{ A/cm}^2,$$

i.e. as long as the emitter current is considerably less than 2 mA. Currents of a few mA, however, are not exceptional, whilst far heavier currents occur with transistors in output-amplifier circuits. In such cases the minority current in the base will no longer be purely a diffusion current, and the influence of the field will have to be taken into account.

*Minority concentrations in the boundary planes of a P-N junction*

The theory put forward in article I was based on the assumption that the minority concentrations in the boundary planes 1 and 2 of a *P-N* junction are fixed by the applied voltage  $V$  at

$$n_p(1) = n_{P0} e^{qV/kT} \text{ and } p_N(2) = p_{N0} e^{qV/kT}, \quad (7a, b)$$

respectively (see I, (1)).

For the sake of convenience we shall introduce in place of  $V$  a new variable  $Z$ , defined as:

$$Z = Q e^{qV/kT}, \dots \dots \dots (8)$$

in which

$$Q = \frac{p_{N0}}{p_{P0}} = \frac{n_{P0}}{n_{N0}} \dots \dots \dots (9)$$

$Q$  is thus the ratio of the equilibrium minority concentration at one side of the *P-N* junction to the equilibrium majority concentration at the other side, and is therefore a very small number. In the transistor chosen as example,  $Q = 2 \times 10^{-7}$ . The equivalence of the two expressions for  $Q$  follows from the condition  $p_{N0}n_{N0} = p_{P0}n_{P0}$  (the product of the equilibrium concentrations of majority and minority charge carriers is independent of the region considered.)  $Z$  is evidently always positive; only if  $V = -\infty$  will  $Z$  become zero. We shall further introduce  $B$ , the ratio of the equilibrium majority concentration in the *P*-region to that in the *N*-region:

$$B = \frac{p_{P0}}{n_{N0}} \dots \dots \dots (10)$$

In a transistor the equilibrium majority concentration in the emitter always substantially exceeds that in the base (cf. I, page 240), so that  $B$  is numerically large. In our sample transistor  $B = 10^4$ .

(7a) and (7b) can now be written in the more convenient form:

$$n_P(1) = n_{N0}Z, \dots \dots \dots (11a)$$

$$p_N(2) = n_{N0}BZ. \dots \dots \dots (11b)$$

In view of our intention to extend the theory to cover higher current densities, these formulae require correction. The values of the four unknown concentrations, viz. the majority and minority concentrations in the two boundary planes of a P-N junction, are given for the general case by four equations which were worked out in an earlier article in this Review <sup>5</sup>). By solving these equations we obtain for the minority concentrations:

$$n_P(1) = n_{N0}Z \frac{1 + BZ}{1 - Z^2}, \dots \dots (12a)$$

$$p_N(2) = n_{N0}Z \frac{B + Z}{1 - Z^2}, \dots \dots (12b)$$

expressions which, for sufficiently small values of Z, reduce to (11a) and (11b). We see from these that both  $n_P(1)$  and  $p_N(2)$  would become infinitely great if Z equalled unity.

According to (8),  $V = (kT/q) \ln(1/Q)$  if  $Z = 1$ . Formula (9) of the article referred to in <sup>4</sup>) shows that this value represents the spontaneous potential difference (contact potential) between the N-region and the P-region. An external voltage compensating the contact potential would thus involve infinitely high minority concentrations and hence an infinitely great current. It will be clear that in practice the external voltages appearing across the P-N junction always lie below the contact potential. In our sample transistor we arrive at a value of 0.3 V for the contact potential (at room temperature  $kT/q = 1/40$  V). The voltages that are applied across the crystal, may, however, be higher than 0.3 V, since the series resistance offered by the crystal itself causes a considerable potential drop.

Z, therefore, is always less than unity. We shall demonstrate in due course that even at a very high current density the emitter voltage is still so low that  $Z \ll 1$ . The terms  $Z^2$  in the denominators of the formulae (12a) and (12b) may therefore be disregarded. Moreover, in view of the high value of B, Z may be neglected in comparison with B in the numerator of (12b), but, for the very same reason, it is not necessarily permissible to neglect

BZ in comparison with unity in formula (12a), so that

$$n_P(1) = n_{N0}Z(1 + BZ), \dots \dots (13a)$$

$$p_N(2) = n_{N0}BZ. \dots \dots \dots (13b)$$

Equation (13b) is identical with (11b), but (13a) can be replaced by (11a) only if  $BZ \ll 1$ , which implies, according to (13b), that  $p_N(2) \ll n_{N0}$ . Here we encounter once more the same condition (see (5)) as the one which makes it possible to treat the minority current in the base as a pure diffusion current. From (13a) and (13b) it follows that, as Z increases from zero (i.e. with increasing voltage in the forward direction and hence increasing current density), the ratio  $n_P(1)/p_N(2)$ , which in the simple theory was considered to remain constant (viz.  $1/B$ ) also increases. This causes, as we shall presently demonstrate, a loss in emitter efficiency.

**The transistor losses as a function of the current density**

The current amplification factor  $\alpha$  is the product of the base efficiency  $\beta$  and the emitter efficiency  $\gamma$  (cf. I, (9)). For the purpose of this article we shall work in terms of losses, namely the total transistor loss  $1 - \alpha$ , the base loss  $1 - \beta$ , and the emitter loss  $1 - \gamma$ . Because  $1 - \beta$  and  $1 - \gamma$ , and hence also  $1 - \alpha$ , are in practice always small with respect to unity, it follows from  $\alpha = \beta\gamma$  that we may write <sup>6</sup>)

$$1 - \alpha = (1 - \beta) + (1 - \gamma). \dots \dots (14)$$

Once we have established the behaviour of  $1 - \beta$  and  $1 - \gamma$  as functions of the current density, we shall know that of  $1 - \alpha$  as a function of the same quantity. Before proceeding to do so, however, we must examine more closely the effect of the electric field on the minority charge carriers in the base of a transistor.

*Apparent increase of the diffusion constant of minority charge carriers owing to the influence of the field*

Field and diffusion drive the minority carriers in the same direction (fig. 3). The effect of the field may thus be regarded as an apparent increase in the diffusion constant. A fact of great importance is that, for very high minority concentrations, the apparent diffusion constant approaches a limit which is twice the actual diffusion constant. We can explain this as follows. The hole current density  $J_p$

<sup>5</sup>) See the article referred to in footnote <sup>4</sup>), equations (12) to (15), in which  $a$  is to be replaced by  $q/kT$  and  $-\Delta\phi$  by  $V$ . Strictly speaking, in formulae (12a) and (12b) of the present article, the factor  $n_{N0}$  should be replaced by  $n_{N0} - p_{N0}$ , whilst B should be written as  $\frac{p_{P0} - n_{P0}}{n_{N0} - p_{N0}}$ . There is, however, no objection to neglecting  $p_{N0}$  and  $n_{P0}$  in comparison with  $n_{N0}$  and  $p_{P0}$  respectively.

<sup>6</sup>) By substituting  $\beta = 1 - (1 - \beta)$  and  $\gamma = 1 - (1 - \gamma)$  in  $\alpha = \beta\gamma$ , and neglecting the very small term  $(1 - \beta)(1 - \gamma)$

and the electron current density  $J_n$ , each written as the sum of the contributions of diffusion and field can be expressed by

$$J_p = -q D_p \frac{dp}{dx} + qp\mu_p F \dots (15a)$$

$$J_n = q D_n \frac{dn}{dx} + qn\mu_n F \dots (15b)$$

respectively ( $\mu_p$  and  $\mu_n$  representing the mobilities of holes and electrons. Apart from this, we have at our disposal the Einstein relation, which expresses the logically acceptable fact that there exists a direct proportionality between the diffusion constant  $D$  and the mobility  $\mu$  for particles having an electric charge of equal magnitude:

$$\frac{D_p}{\mu_p} = \frac{D_n}{\mu_n} = \frac{kT}{q} \dots (16)$$

Finally we know from the neutrality condition (valid for regions outside a  $P-N$  junction) that the concentration gradients for holes and electrons are equal:

$$\frac{dn}{dx} = \frac{dp}{dx} \dots (17)$$

By eliminating  $F$  from (15a) and (15b) and using (16) and (17), we obtain for  $J_p$ :

$$J_p = -q \frac{1 + \frac{P}{n}}{1 - \frac{J_n \mu_p P}{J_p \mu_n n}} D_p \frac{dp}{dx} = -q D_p' \frac{dp}{dx} \dots (18)$$

This expression for  $J_p$  has the form applicable to a diffusion current density. The influence of the field is expressed in the substitution of the apparent diffusion constant  $D_p'$  for the ordinary diffusion constant  $D_p$ . In an  $N$ -region, where we write  $p_N$  instead of  $p$ , we find after substitution from  $\mu_p/\mu_n = \frac{1}{2}$  and from  $n_N = p_N + n_{N0}$  (neutrality condition, see (4)), that

$$D_p' = \frac{1 + \frac{P_N}{P_N + n_{N0}}}{1 - \frac{J_n P_N}{2J_p P_N + n_{N0}}} D_p \dots (19a)$$

As the current in the base of a transistor is mainly composed of holes,  $J_n/J_p \ll 1$ , allowing us to write:

$$D_p' = \left(1 + \frac{P_N}{P_N + n_{N0}}\right) D_p \dots (19b)$$

this changing for very high values of the minority concentration  $p_N$  into:

$$D_p' = 2 D_p \dots (20)$$

Fig. 4 shows how  $D_p'/D_p$  gradually increases from 1 to 2 as the value of  $p_N/n_{N0}$  increases. The apparent increase in the diffusion constant is noticeable at high current densities, since these are accompanied by high minority concentrations in the base.

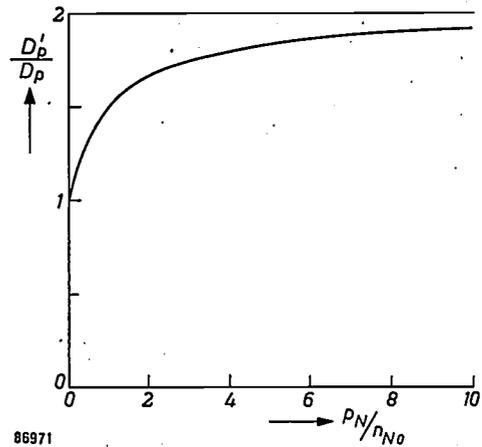


Fig. 4. The ratio  $D_p'/D_p$  of the apparent to the actual diffusion constant as a function of the ratio  $p_N/n_{N0}$  of the minority concentration to the equilibrium majority concentration in the base of a  $P-N-P$  transistor.

*Reduction of base losses at high current densities*

Owing to the effect of the field in the base of a  $P-N-P$  transistor at high current densities, the holes are transported more quickly through the base and thus run less risk of loss through recombination. The base loss, therefore, will diminish at increasing current density. The hole concentration is always highest near the emitter, in cross-section 2 (fig. 3). Here it will be most noticeable that the apparent diffusion constant  $D_p'$  deviates from the normal value  $D_p$  and it is here too that  $D_p'$  will come nearest to its limit of  $2 D_p$ . At very high current densities (theoretically only at an infinitely high value)  $D_p'$  will reach the value  $2 D_p$  throughout the whole base (i.e. up to infinitely near the collector). The average time a hole spends in crossing the base will then be halved. If we assume that the average lifetime  $\tau_b$  of a hole in the base is independent of current density, then that proportion of the total number of holes arriving per second in the base lost by recombination, will also be halved. Hence, at increasing current density, the base loss approaches half the value it has at low current densities. Upon further analysis (into which we

shall not enter here) it is found that the base loss can be represented to a close approximation by simply substituting  $D_p'$  for  $D_p$  in the expression (I, 12). If in addition  $D_p'$  is replaced by the expression (19b), we find that:

$$1 - \beta = \frac{w^2}{2 \tau_b D_p} \frac{1 + p_N(2)/n_{N0}}{1 + 2p_N(2)/n_{N0}} \dots (21).$$

In fig. 5  $1-\beta$  is plotted as a function of  $p_N(2)/n_{N0}$ , on the basis of the numerical values given in our example.

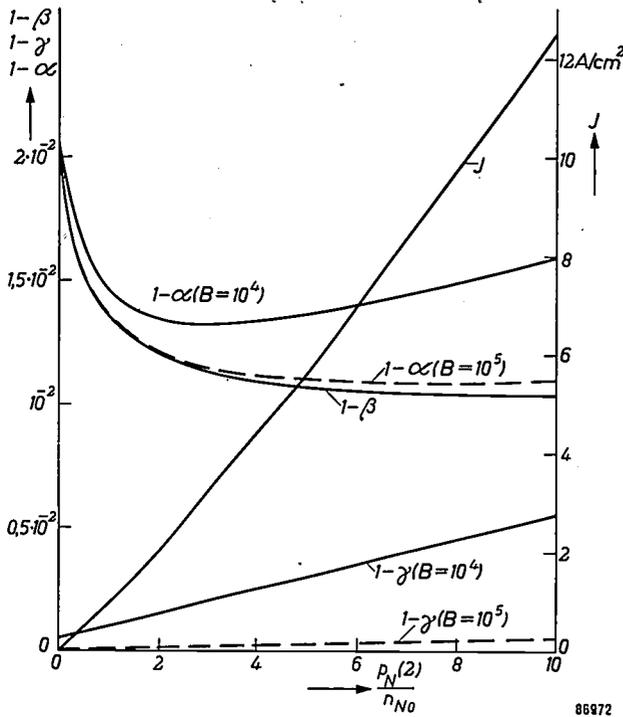


Fig. 5. Base loss  $1-\beta$  ( $\beta$  = base efficiency), emitter loss  $1-\gamma$  ( $\gamma$  = emitter efficiency) and current density  $J$  as functions of  $p_N(2)$  (minority concentration in the boundary plane of the base at the emitter side), using  $n_{N0}$ , the equilibrium majority concentration in the base, as the unit. The full curves apply to the transistor chosen as an example in this article, for which  $B (= p_{P0}/n_{N0}) = 10^4$ ; the dotted curves refer to a transistor whose equilibrium concentrations in base and emitter have been so modified that  $B = 10^5$ .

The premise that  $\tau_b$ , the average lifetime of the holes in the base, is independent of the current density, i.e. independent of the concentrations of holes and electrons, even when the concentrations are very high, is not an immediately obvious one. In such circumstances every hole encounters a very large number of electrons, and one would expect the chance of recombination to increase for every hole, with the consequence that  $\tau_b$  is shortened. In fact, however, recombination is a very intricate process. Direct recombinations occur very rarely; as a rule the process requires a number of intermediate steps, so that  $\tau_b$  may well remain virtually constant. The

assumption that it does so is supported by the results of measurements.

In the base of a  $P-N-P$  transistor the hole concentration gradually decreases towards zero from emitter to collector, so that according to (19b) the (apparent) diffusion constant  $D_p'$  gradually decreases towards  $D_p$ . Disregarding generation and recombination in the base, we may consider  $J_p$  to be independent of the position  $x$ , and according to (18),  $dp/dx$  must then gradually increase from emitter towards collector. The concentration curve of the holes, therefore, is no longer linear. The calculations of base and emitter efficiency in I, however, were based on a supposedly nearly linear concentration curve. This concentration curve can also be calculated at high current densities by integration of (18), after substituting the expression (19b) for  $D_p'$ . The boundary conditions are that the values of  $p$  at emitter and collector must be  $p_N(2)$  and 0 respectively. We obtain in this way the curves shown in fig. 6. For very small, but equally so for very large values of  $p_N(2)$ , the deviation from linearity is negligible. In the latter case  $p$  is very large up to the direct vicinity of the collector, so that over almost the whole of the base  $D_p'$  has nearly the constant value of  $2 D_p$ . Hence  $dp/dx$  is also constant, according to (18). Only in the immediate vicinity of the collector will the gradient increase, doubling its value at the collector itself (curve d). A maximum deviation from linearity must then occur between the extreme cases of very low and infinitely high values of  $p_N(2)$ . That this maximum deviation is still fairly small is apparent from the form of curve c. In our example this case occurs at a current of about 8 mA. The approximate formulae (21) and (22) for the base and emitter losses are not affected by this slight deviation from linearity.

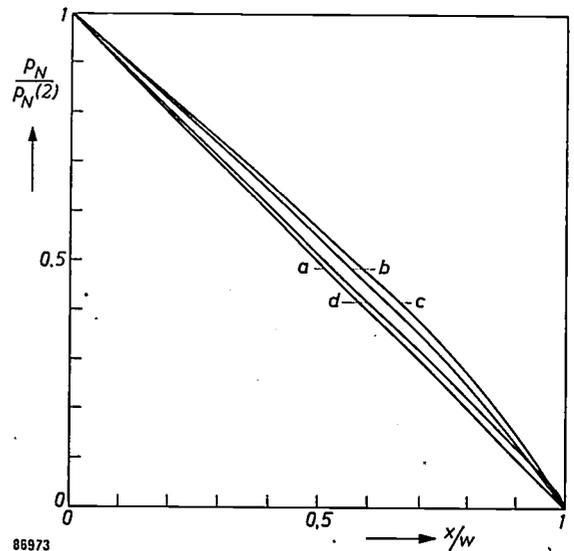


Fig. 6. The hole concentration  $p_N$  in the base of a  $P-N-P$  transistor shows a spatial variation that is always nearly linear. The curves shown here are plotted for different values of  $p_N(2)/n_{N0}$ . The origin,  $x = 0$  lies at cross-section 2, at the emitter. To facilitate comparison of the curves as regards their linearity, the quantity actually plotted is the ratio  $p_N/p_N(2)$ , whereby all curves start in point 1 on the ordinate. Curve a:  $p_N(2)/n_{N0} \ll 1$ ; curve b:  $p_N(2)/n_{N0} \approx 1$ ; curve c:  $p_N(2)/n_{N0} \approx 2$  (greatest deviation from linearity). For still greater values of  $p_N(2)/n_{N0}$  (curve d) the curve approaches a again.

*The increase in the emitter loss at high current densities*

We should prefer the total current through the emitter junction to be carried entirely by holes, but in fact a minor part of it is composed of electrons. In the emitter the electron current is a pure diffusion current, so that  $J_n$  is proportional to  $n_p(1)$ <sup>7)</sup>. If the hole current in the base were also a pure diffusion current, then  $J_p$  would be proportional to  $p_N(2)$  and  $J_n/J_p$  would be proportional to  $n_p(1)/n_N(2)$  in consequence. However, as we have shown on page 64, this ratio becomes higher as  $J$  increases, so that an increasing portion of  $J$  would be composed of electrons. There is a certain amount of compensation owing to the fact, likewise discussed earlier, that the electric field comes to the aid of the hole current, so that initially  $J_p$  increases (proportionately) more rapidly than  $p_N(2)$ . Upon further analysis, however, it turns out that this effect of the field cannot fully compensate the effect of the increase of  $n_p(1)/p_N(2)$ , so that  $J_n/J_p$  nevertheless always increases with  $J$ . The increase is accompanied by a falling-off of the emitter efficiency  $\gamma$  and hence with an increase in emitter loss  $1-\gamma$ . The analysis shows that the expression for  $1-\gamma$ , which could be derived from (I, 11), requires a correction factor  $(1+p_N(2)/n_{N0})$ . We then arrive at

$$1-\gamma = \frac{D_n w}{D_p L_e B} \left( 1 + \frac{p_N(2)}{n_{N0}} \right), \dots (22)$$

in which  $B$  again is given by (10).  $L_e$  represents the diffusion length in the emitter (fig. 3). The fully drawn curve in fig. 5 applies to  $1-\gamma$  plotted for our example<sup>8)</sup>.

The total loss  $1-\alpha$ , consisting according to (14) of the sum of emitter and base losses, is also plotted in fig. 5 as a function of  $p_N(2)/n_{N0}$ . This diagram further shows the current density  $J(=J_p)$ , and from this we may derive a curve showing  $1-\alpha$  as a function of  $J$ . The full curve of fig. 7 was obtained in this purely theoretical manner; it does in fact have the same character as the experimentally derived curve of fig. 2.

It appears from the curve for  $J$  in fig. 5 that  $J$  is already very large when  $p_N(2)/n_{N0} = 10$ . It

7) See I, p. 236, eq. 4. Strictly speaking,  $J_n$  is proportional to  $n_p(1)-n_{p0}$ , but  $n_{p0}$  can in practice always be neglected in comparison with  $n_p(1)$ .

8) The concentration change of the electrons in the emitter as shown in fig. 3 is not quite as it is in fact. The thickness  $d_e$  of the emitter is very small (e.g. 20  $\mu$ ). At the point of contact with the electrode, equilibrium concentrations occur. Owing to this, the concentration gradient of the electrons in cross-section  $l$  is not determined by  $L_e$ , but by  $d_e$ . Consequently, we have put  $L_e = 20 \mu$  in the evaluation of formula (22).

follows from (13b) that in this case  $BZ = 10$  so that, with  $B = 10^4$ ,  $Z = 10^{-3}$ . Neglecting  $Z^2$  with respect to 1 and  $Z$  with respect to  $B$ , as was done in the derivation of (13a) and (13b), was, therefore, perfectly justified.

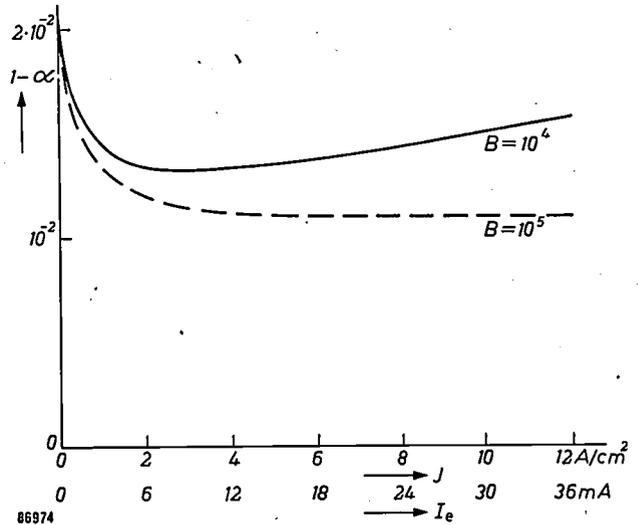


Fig. 7. The losses  $1-\alpha$  of a transistor as a function of the current density  $J$ , as derived from fig. 5, for two values of  $B$ . A scale of emitter current values ( $I_e$ ) based on an emitter surface of 0.3 mm<sup>2</sup>, is also set out along the abscissa.

**Reduction of transistor losses**

Formulae (21) and (22) show how various quantities influence the transistor losses. The influence of the factor  $B$ , which is defined by (10), is of particular importance. The base loss is evidently independent of  $B$ , whereas  $1-\gamma$  is inversely proportional to it. If, for example,  $B$  is multiplied by a factor of 10, the slope of the curve for  $1-\gamma$  will become 10 times smaller. Only for much larger values of  $p_N(2)$ , then, will  $1-\alpha$  (dotted curve in fig. 5) diverge perceptibly from  $1-\beta$ . Analysis shows that for  $J$  we can write:

$$J = q \frac{D_p n_{N0}}{w} \left\{ 2 \frac{p_N(2)}{n_{N0}} - \ln \left( 1 + \frac{p_N(2)}{n_{N0}} \right) \right\},$$

an expression in which  $B$  does not appear, so that the curve  $J$  in fig. 5 is independent of the value of  $B$ . Thanks to the larger value of  $B$ , the curves  $1-\gamma$  and  $1-\alpha$  as functions of  $J$  have the much more favourable form shown as a dotted curve in fig. 7.

In order to raise the value of  $B (=p_{P0}/n_{N0})$ , we may attempt to increase  $p_{P0}$ , the concentration of the holes in the emitter. This can be done by incorporating a large number of acceptor atoms in the germanium of the emitter<sup>9)</sup>. For the provision of acceptor atoms, indium is often used, a substance

9) See the article referred to in footnote 4), especially fig. 2.

which offers certain advantages in the manufacturing of "alloy transistors". As regards its solubility in germanium, however, another acceptor element such as gallium seems preferable, in view of the fact that the highest  $p_{P0}$  attainable with indium is approximately  $5 \times 10^{18} \text{ cm}^{-3}$ ; whereas with gallium which is much more soluble in germanium, a concentration ten times higher can be attained. As against this, gallium has certain properties which render it unsuitable as a direct substitute for indium. In our Eindhoven laboratories experiments have been carried out using an acceptor material consisting of indium in which some gallium had been dissolved. In this way, gallium can be dissolved in the germanium together with the indium with the result that  $p_{P0}$ , and hence the factor  $B$ , can be made 10 times greater. The advantages of indium from the manufacturing point of view are thus combined with the superior solubility of gallium in germanium.

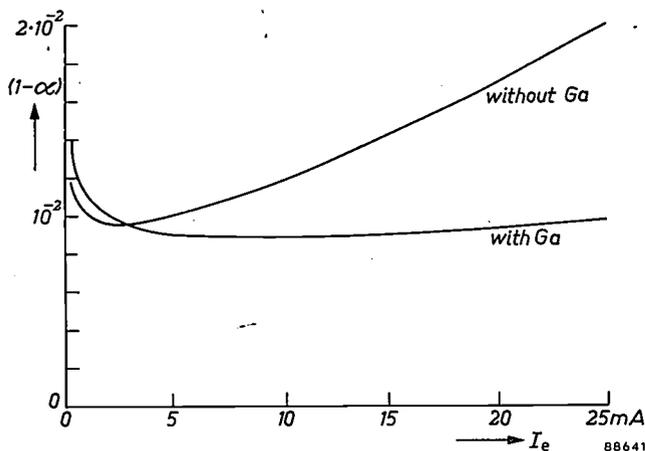


Fig. 8. The transistor losses as a function of the emitter current for a transistor whose emitter contains gallium in addition to indium. For comparison the curve from fig. 2 is also shown, which refers to a similar transistor containing no gallium.

Favourable results have indeed been achieved by this process, as appears from fig. 8, in which  $1-\alpha$  is plotted as a function of the emitter current  $I_e$  in an experimental transistor built on this principle.

A higher equilibrium concentration of majority charge carriers in the emitter is a particularly attractive feature in power transistors (transistors with a relatively large power-handling capacity), in which current densities are always high, but the process can also be profitably applied in the manufacture of transistors not intended for high current densities. Raising  $p_{P0}$  allows  $n_{N0}$  to be raised while keeping the same values for  $B$  and hence for  $1-\gamma$  (see (23)). This means that a base material of higher conductivity can be chosen (since the conductivity is higher according as the majority equilibrium concentration — and hence  $n_{N0}$  in a base of  $N$  material — is higher). This is of advantage for limiting the influence of the "Early effect" (discussed in I, page 245). The same principle can also be applied in the manufacture of transistors for high frequencies, which will be discussed in a subsequent article.

**Summary.** At high current densities the simple theory of the junction transistor is no longer valid, in view of the fact that in the base the minority concentration ceases to be everywhere negligible with respect to the majority concentration. One consequence of this is that the base loss drops fairly quickly to half its initial value as the current density is raised; another consequence is that the emitter loss gradually increases. The increase in emitter loss continues long after the base loss has become constant, the nett result being that the total losses go on increasing and that the current amplification factor falls off. The higher the equilibrium concentration of majority charge carriers in the emitter, the slower the rise of emitter losses with current density; as a result of this the excessive losses are not incurred until higher current densities. Where  $P-N-P$  alloy transistors are concerned, this higher concentration can be obtained by the employment of a gallium-indium alloy as acceptor material for the emitter instead of the usual pure indium.

## ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk \* can be obtained free of charge upon application to the Philips Research laboratory, Eindhoven, Netherlands.

**2281:** J. L. Meijering: On the diffusion of oxygen through solid iron (Acta Met. 3, 157-162, 1955, No. 2).

Internal oxidation of three iron alloys showed — at least at  $1200^\circ \text{C}$  — no trace of intergranular preference. Therefore, in contrast to a conjecture of Kitchener et al., normal interstitial diffusion of oxygen through the iron lattice must be assumed. Approximate values of the permeability (product of diffusion coefficient and solubility) of O in Fe at different temperatures are calculated from the

sub-scale thickness. The permeabilities of Fe, Ni, Cu and Ag for oxygen at 85% of their absolute melting temperatures are of the same order of magnitude. Discrepancies in the solubilities of O in solid Fe found by different authors are discussed.

**2282:** C. J. Bakker: Production and physical properties of radio-isotopes (Rev. Trav. chim. Pays-Bas 74, 281-293, 1955, No. 5).

The production of radio-isotopes in nuclear reactors and in cyclotrons is discussed. The types

of nuclear reactions involved are mentioned. As to the production of radio-isotopes, the cyclotron is an important complement to the nuclear reactor. Examples of disintegration schemes of radio-isotopes are given. The loss of energy of  $\beta$  or  $\gamma$ -rays during their passage through matter is briefly discussed. The precautions necessary when working with radio-isotopes are mentioned.

**2283:** A. M. J. Jaspers and H. R. Marcuse: Local measurement of  $I^{131}$  uptake in the thyroid gland (*Rec. Trav. chim. Pays-Bas* 74, 355-361, 1955, No. 5).

Based on a new development of Geiger-Müller tubes, a collar-shaped device is described, which can be applied round the patient's neck in order to measure directly the uptake of  $I^{131}$  in the thyroid gland. Methods are indicated for calibrating and for measuring the corrections due to the background and the activity in the blood.

**2284:** J. A. M. Dikhoff and N. W. H. Addink: Ergebnisse von Intensitätsbestimmungen mit der s.p.d. Skala (*Microchim. Acta*, pp. 257-264, 1955, No. 2-3). (Results of intensity measurements using the standard paper density scale; in German).

Addink has described a simple method for the visual determination of line intensities by means of the so-called standard paper density scale. In the present paper closer attention is given to the background correction, the calibration of the scale, the accuracy of the method and some other factors which may influence the measurements. Finally a short discussion is given of scale lines having a profile similar to that of spectrum lines.

**2285:** N. W. H. Addink: Leitprobenfreie Verfahren (*Microchim. Acta*, pp. 703-708, 1955, No. 2-3). (Methods of spectrochemical analysis without standard samples; in German).

It is shown that in the addition method of spectrochemical analysis (in which the unknown concentration of an element is obtained by extrapolation) it is still necessary to check that the intensity-concentration curve is linear when plotted with logarithmic coordinates. Harvey's method is considered, in which the concentration is determined from the ratio of the intensity of a spectral line to the background. In this connection, the intensity of the background may not be used as an "internal standard". Finally details are tabulated of methods of analysis which do not use comparison samples.

**2286:** N. W. H. Addink and L. J. P. Frank: Zinc in relation to cancer (*Naturwiss.* 42, 419-420, 1955, No. 14).

It is demonstrated that fresh venous blood of cancer patients shows a lower zinc concentration than that of healthy people. As, moreover, the zinc content of tumours is lower (about one third) than that of healthy tissues, the zinc content of the blood provides a picture of the tumour.

Examples are given showing favourable and unfavourable progress of the disorder. In the favourable cases the zinc concentration of the blood was gradually increasing with time, whereas in the unfavourable (fatal) ones low values were found. A conjecture concerning the natural and artificial development of the disease is mentioned.

**2287:** H. C. Hamaker: Naar efficiënte experimenten (*Statistica neerlandica* 9, 7-25, 1955, No. 1-2). (Efficient experimentation; in Dutch).

Under industrial conditions which involve large random fluctuations, large numbers of observations and many influencing factors, analysis of data on a statistical basis has become an essential part of experimental techniques, which ultimately leads to more efficient experimentation. Seven examples are given. The first four are simple experiments for the comparison of two batches when disturbing factors such as a time-trend must be eliminated. The other three examples discuss more complex cases, viz. a 3-factor experiment in a foundry, a latin square experiment in nylon spinning and a five-factor experiment of a mixed kind on gear-noise testing. An attempt is made to show in a simple form the relation between the statistical analysis and the conclusions that may be drawn from the results of a well-designed experiment.

**2288:** G. Diemer: Electroluminescentie (*Ned. T. Natuurkunde* 21, 165-176, 1955, No. 7). (Electroluminescence; in Dutch).

Survey of existing theory and experiments on electroluminescence, including the theories of Piper and Williams, of Alfrey & Taylor and of Zalm, Diemer and Klasens. Possible applications of the phenomenon are mentioned.

**2289:** J. D. Fast and M. B. Verrijp: Solubility of nitrogen in alpha-iron (*J. Iron Steel Inst.* 180, 337-343, 1955).

The solubility of nitrogen in  $\alpha$ -iron as a function of temperature can be determined accurately by measuring the internal friction of nitrogen-containing iron wires. The value of the internal friction is

proportional to the amount of dissolved nitrogen right up to saturation ( $\sim 0.1\%$  by weight). Three different solubilities can be distinguished, viz. those in which the iron is in equilibrium with "Fe<sub>8</sub>N", with Fe<sub>4</sub>N and with N<sub>2</sub> at atmospheric pressure. All three solubilities were determined. "Fe<sub>8</sub>N" is a metastable nitride which forms only below 250 °C from supersaturated solutions. Both "Fe<sub>8</sub>N" and Fe<sub>4</sub>N form only because of the slowness of the reaction at the outer surface in which N<sub>2</sub> molecules are formed from dissolved N atoms. If this reaction were not slow, the equilibrium between  $\alpha$ -iron and "Fe<sub>8</sub>N" and Fe<sub>4</sub>N could be studied only under large N<sub>2</sub> pressures. These pressures, corresponding to the dissociation pressures of the nitrides can be calculated from the measured solubilities. The results agree satisfactorily with the measurements of American investigators.

**2290:** H. B. G. Casimir: Ferro- und Antiferromagnetismus (Physikertagung Hamburg 1954, Hauptvorträge, 139-152). (Ferromagnetism and anti-ferromagnetism; in German).

Survey, from a theoretical standpoint, of various aspects of ferromagnetism and anti-ferromagnetism for physicists and electrical engineers not specialized in this field. Two groups of phenomenon are distinguished: those involving the magnetization within a single Weiss-domain and those involving the combined magnetization over a large number of domains. Of the first group, the following are dealt with: saturation magnetization, Curie temperature, the internal magnetic field which results from the exchange force theory of Heisenberg, crystal and strain anisotropy, gyromagnetic resonance and the introduction by Néel of "antiferromagnetism" to explain certain phenomena in ferrites. Néel's view that the magnetization in ferrites is due to two opposed and incompletely compensated magnetized systems (uncompensated antiferromagnetism) is confirmed by experimental work by Gorter at Philips Research Laboratories, Eindhoven.

The phenomena of the second group consists of macroscopic magnetization by the displacement of Bloch walls and by rotation of the magnetization within Weiss domains. In soft-magnetic alloys, wall displacements are at first responsible for the magnetization — and thus determine the initial permeability — while rotation processes appear only at higher field strengths. In ferrites, there are strong indications that the situation is precisely the reverse. This view is supported by measurements of the A.C. magnetization of ferrites as a function of amplitude and frequency.

The article is concluded with some remarks on hard-magnetic materials.

**2291:** J. S. C. Wessels: Photo-oxidation of ascorbic acid by isolated chloroplasts (Rec. Trav. chim. Pays-Bas, 74, 832-840, 1955, No. 7).

The photo-oxidation of ascorbic acid by chloroplasts in the presence of various quinones and dyes was investigated. The kinetics of the reactions were studied, using 2,6-dichlorophenol-indophenol as dye mediator. Inhibitors of the Hill reaction did not affect the photo-oxidation rate. It is postulated that the reaction is photosensitized by chlorophyll. Evidence for this assumption was obtained by bringing about quite the same reaction in alcoholic solution, the chloroplasts being replaced by chlorophyll a. A possible role of this photo-oxidation in the back-oxidation of reduced Hill-reaction oxidants is discussed.

**2292:** D. J. Braak: New mobile and auxiliary equipment produced by Philips Telecommunication Industries (T. Ned. Radiogenootschap 20, 161-174, 1955, No. 3).

A new portable trans-receiver has been developed with five channels. Two types are normally made, viz. for the 80 Mc/s band and for the 160 Mc/s band but a third is also made to order for the 40 Ms/c band. After a general description of the equipment and its construction some of the more important auxiliary equipment is briefly described.

**2293:** L. R. Bourgonjon: Circuits for frequency generators (T. Ned. Radiogenootschap 20, 175-181, 1955, No. 3).

Present day intensive use of radio equipment calls for higher demands on frequency stability. When high stability must be combined with variation in frequency, it is necessary to use a system for frequency synthesis. A survey of two different systems for frequency synthesis is given, one system using a mixer stage, the other employing a control loop with phase discriminator. To economize on crystals, in both systems an impulse governed oscillator may be used. The block diagram of a multi-channel transmitter-receiver in the 100-156 Mc/s frequency band is shown.

**2294:** H. C. Bennebroek Evertsz: Radio link operating in 4000 Mc/s band for television and multi-channel telephony (T. Ned. Radiogenootschap 20, 183-192, 1955, No. 3).

A radio link equipment of new design is described. It comprises heterodyne repeater stations and is equipped with S.H.F. triode amplifiers.

**2295:** W. L. Verwest: Automatic tuning of transmitters (T. Ned. Radiogenootschap 20, 195-206, 1955, No. 3).

The Philips Instantuner is a device for automatically resetting a tuning element in any one of twelve preset positions. When several tuning elements are used, which is normally the case, a corresponding number of Instantuners is required. Remote control of radio equipment is then limited to a simple manipulation. Manual tuning is also possible in any position by pressing a push-button mounted in a special vernier knob. The presetting procedure is very easy. Changing from one preset frequency to another (automatic tuning) is accomplished in 1-3 seconds for small transmitters, e.g. aircraft transmitters, and in 2-10 seconds for larger transmitters, e.g. broadcast transmitters. Each Instantuner is built as a separate unit, designed for universal application. It consists of a blocking-mechanism and a torque-limiting clutch. The blocking-mechanism contains twelve pawls and pawl rings, so that the main shaft can be blocked in any one of twelve preset positions. Owing to the use of the special torque-limiting clutch a resetting accuracy of  $\pm 0.01^\circ$  can be achieved. The clutch releases the driven spindle of a tuning element from the driving-motor, as soon as the preset position is reached.

**R 281:** W. L. Wanmaker and M. L. Verheyke: A thermogravimetric study on the preparation of calcium halophosphate (Philips Res. Rep. 11, 1-18, 1956, No. 1).

A thermogravimetric analysis is carried out on the raw materials used in the preparation of the calcium halophosphate phosphor activated by Sb and Mn. Furthermore, some binary and ternary mixtures of these raw materials and some mixtures producing halophosphate on firing are investigated. The dissociation temperature of  $\text{CaCO}_3$  is depressed by the presence of  $\text{CaHPO}_4$  and  $\text{CaHPO}_4 + \text{SrCl}_2$ . The dissociation of  $\text{CaCO}_3$  is facilitated to a still greater degree in mixtures producing halophosphates on firing. The dissociation temperature of  $\text{CaCO}_3$  rises in mixtures producing fluorochloro-, chloro- and fluoroapatite (in the order given). The analysis of halophosphate firing mixtures shows that the loss in Sb, Cl and F occurs only after complete dissociation of the  $\text{CaCO}_3$ . The loss of these elements is high with a low M/6  $\text{PO}_4$  ratio ( $M = \text{Ca} + \text{Sr} + \text{Mn}$ ). The loss in chlorine may occur as a result of pyrohydrolysis of  $\text{SrCl}_2$  or of the formation of  $\text{SbCl}_3$ . It is quite probable that the thermobalance could also be used with success for the study

of other solid-state reactions in which a loss in weight occurs.

**R 282:** J. van Slooten: Oscillations and noise (Philips Res. Rep. 11, 19-26, 1956, No. 1).

In the first section it is shown that there is no well-defined boundary between "order" and "disorder". From a disorderly distribution in time an orderly one may be derived by mathematical or technical means. It is not surprising, therefore, that oscillations and noise may be considered from a common point of view. In the second section the similarity and the difference between a narrow noise band and an oscillator are treated in some detail. Subsequently the phase and frequency fluctuations of an LC oscillator are studied by means of the method of phase jumps.

**R 283:** J. van den Boomgaard: Zone-melting processes for pure compounds AB with a negligible vapour pressure (Philips Res. Rep. 11, 27-44, 1956, No. 1).

The composition of a pure compound AB generally will change if zone-melting processes are applied to it. In the present paper the theory for this case is developed under the assumption that reactions with the vapour may be neglected, i.e. the compound has no measurable vapour pressure. With a few general assumptions, which will be clarified by a model describing the solid and the liquid state of binary semiconductors in terms of vacancies, it is possible to define the distribution equilibria with two distribution constants and to derive mathematical expressions for the deviation from the stoichiometric compositions as a function of the place in the rod.

**R 284:** A. H. Boerdijk: Technical aspects of levitation (Philips Res. Rep. 11, 45-56, 1956, No. 1).

Levitation of a body is here defined as a state of either stable or neutral equilibrium relative to the earth, in which material contact between the body and its environment is not essential. The possibilities and limitations of levitation by auxiliary gravitational forces, by reaction forces and by forces in electromagnetic fields are investigated. Levitation by gravitational forces or by radiation pressure is not feasible in practice, whilst levitation by forces in electrostatic fields is theoretically impossible. Under certain conditions levitation may be achieved by reaction forces and by forces in magnetostatic, stationary and quasi-stationary electromagnetic fields. Published applications comprise balances

centrifuges, and a method for melting metals in vacuo without a crucible.

**R 285:** K. F. Niessen: Condition for resonance in a nearly compensated ferrimagnetic (Philips Res. Rep. **11**, 57-65, 1956, No. 1).

The resonance condition is derived for a ferrimagnetic with two sub-lattices with different total moments, different anisotropy constants and different gyromagnetic ratios. These three differences are assumed to be relatively small. The above result may be applied, for example, to an antiferromagnetic in whose sub-lattices a relatively small number of magnetic ions has been replaced by non-magnetic ones.

**R 286:** B. van der Veen: The equivalent network of a piezoelectric crystal with divided electrodes (Philips Res. Rep. **11**, 66-79, 1956, No. 1).

A piezoelectric-crystal plate with divided electrodes can be considered as a four-pole. A theoretical derivation is given of the equivalent electrical circuit of this four-pole for frequencies close to the mechanical resonant frequency of the plate. The four-pole is compared with the equivalent circuit as put forward by Mason, Sykes, Herzog, and others.

**R 287:** J. Volger and J. M. Stevels: Electric polarizability of colour centres in quartz crystals and glasses (Philips Res. Rep. **11**, 79-80, 1956, No. 1).

High energy radiation can create colour centres in quartz which give rise to dielectric losses (see these abstracts **R 272**). A possible mechanism to explain these losses is suggested and discussed in connection with various models for colour centres put forward by a number of workers.

**R 288:** A. van Weel: Error sources in group-delay measurements on electric networks (Philips Res. Rep. **11**, 81-90, 1956, No. 2).

Errors in group-delay measurements can be caused by spurious phase modulation in the amplitude modulator or in the network under test, by a varying carrier-frequency level on the detector, and by overloading the network under test. These effects are discussed and counter measures are indicated.

**R 289:** J. van den Boomgaard: Zone-melting processes for compounds AB with a measurable vapour pressure under influence of the atmosphere (Philips Res. Rep. **11**, 91-102, 1956, No. 2).

Zone melting of compounds AB, in which deviations from the exact stoichiometric composition may occur and which have a measurable vapour pressure at the melting point, is possible if a vapour of one of the components is applied to prevent the decomposition of the material. If the liquid AB is described in a semi-crystalline way with two kinds of coupled defects, which are in equilibrium with two kinds of coupled defects in the solid, it is possible to describe the distribution equilibrium and the reaction with the gas phase in terms of these defects. With this model, it is possible to derive some mathematical expressions for the relation between the deviation from the exact stoichiometric composition and the position in the rod of AB, as a result of zone-melting processes.

**R 290:** J. M. Stevels and A. Kats: The systematics of imperfections in silicon-oxygen networks (Philips Res. Rep. **11**, 103-114, 1956, No. 2).

The structure of the  $\text{SiO}_2$  network permits a wide variety of possible imperfections, a number of which are discussed. A system is developed by which these imperfections may be described. The system is then extended so as to allow of the description of the structure of glasses. Besides this "synthetic" system for the nomenclature of the centres, an "analytical" system is developed. The advantages of both systems are discussed.

**R 291:** A. Kats and J. M. Stevels: The effect of U.V. X-ray radiation on silicate glasses, fused silica and quartz crystals. (Philips Res. Rep. **11**, 115-156, 1956, No. 2).

Glasses, fused silica and quartz crystals are compounds with open structures which favour the occurrence of interstitial cations and anions. A systematic investigation is made of the effect of ultraviolet radiation, X-rays and electrons. The centres thus formed are discussed in terms of the new nomenclature developed recently for open structures. Generally speaking, glasses and quartz crystals behave similarly, whereas fused silica has many features in common with amethyst.

# Philips Technical Review

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

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

## A PENTODE GUN FOR TELEVISION PICTURE TUBES

by J. C. FRANCKEN, J. de GIER and W. F. NIENHUIS.

621.385.832.032.269.1: 621.397.62

*As a result of improvements in the quality of the transmitted television picture, higher demands are being made on picture tubes. The characteristics of the electron gun play a fundamental part in this respect. Owing to the use of different line systems in different countries, it can be desirable, as this article will show, to have a gun that is also capable of variation. With the "pentode" gun described here the optimum quality can be simply adjusted for any line system.*

The function of the electron gun in a television picture tube is to form a beam of electrons. By means of an electron lens, the electron beam is then focussed upon the fluorescent screen where it appears as a spot of light. The spot of light scans a frame the brightness of which must vary synchronously point by point with the intensity of the received television signal. For this purpose the electron gun is provided with a control grid, which controls the current of the electron beam as a function of the signal voltage.

The simplest type of electron gun (see *fig. 1*) consists of three electrodes (triode gun): the cathode ( $k$ ), the control grid ( $g_1$ ) and a second electrode ( $g_2$ ) usually called the anode, which gives the electrons the energy needed to produce a bright picture. The electric field between  $g_2$  and  $g_1$  functions as a positive lens and thus makes the beam less divergent.

The characteristics of a triode gun are closely related to those of a triode valve. For example, in a triode gun the current  $I_s$  flowing to the anode — and through the anode to the screen — is highly dependent upon the anode voltage  $V_{g_2}$ . At moderate currents, the brilliance of the image on the screen is proportional to the product  $I_s V_{g_2}$ . As a result, the brilliance varies more than proportionally to  $V_{g_2}$ . This is usually undesirable, in view of the difficulty and expense involved in adequately stabilizing the anode voltage  $V_{g_2}$ , which may be between 12 and 16 kV in modern tubes.

For this reason the electron gun most commonly used is the tetrode type as sketched in *fig. 2*. The anode in this gun consists of two parts: the first and final anodes  $g_2$  and  $g_3$ . The first anode carries a potential of 300 V and is so designed that hardly

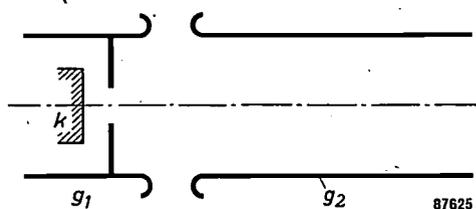


Fig. 1. The simplest form of electron gun (triode gun).  $k$  cathode,  $g_1$  control grid,  $g_2$  anode.

any electrons fall on it. There is no difficulty in stabilizing this voltage. The beam current  $I_s$  is determined almost entirely by  $V_{g_1}$  and  $V_{g_2}$ , and is practically unaffected by the high potential  $V_{g_3}$  on the final anode. The brilliance is now only

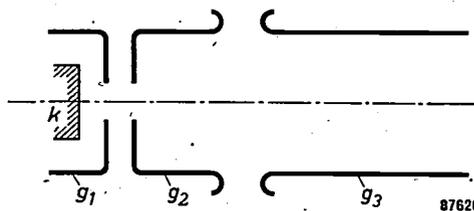


Fig. 2. Tetrode gun.  $k$  cathode,  $g_1$  control grid,  $g_2$  first anode,  $g_3$  final anode.

proportional to the final anode voltage (here  $V_{g_3}$ ), so that this voltage need no longer be so accurately stabilized. Another important advantage of the tetrode gun is that, owing to the lens action of the field between  $g_2$  and  $g_3$ , the emergent beam is more convergent than is the case on a triode (more prefocussing). Moreover, the extra parameter allows more latitude in selecting the degree of prefocussing required. As will be shown, this characteristic has an important bearing on the quality of the picture.

Fig. 3 represents schematically the complete electrode system of a picture tube, showing the path of the electrons. One may regard the system as consisting of four components: the gun with

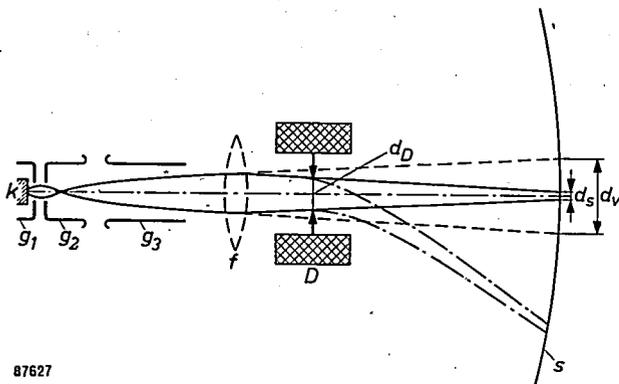


Fig. 3. The electrodes of a picture tube and the path of the electron beam.  $K$ ,  $g_1$ ,  $g_2$  and  $g_3$  together constitute the tetrode gun.  $D$  deflection coils,  $s$  screen,  $d_v$  and  $d_s$  diameters of focussed and unfocussed spots respectively,  $d_D$  beam width of focussed beam in the deflection coils.

electrodes  $k$ ,  $g_1$ ,  $g_2$  and  $g_3$ , the focussing lens  $f$ , the deflection coils  $D$  and the screen  $s$ . Excluding the action of the lens  $f$ , the electron gun prefocusses the beam to produce a spot of light of diameter  $d_v$  on the screen. If the lens  $f$  is now introduced, the beam is focussed sharply, resulting in a smaller spot of diameter  $d_s$ . The width of the beam in the deflection coils is then  $d_D$ . By means of the magnetic field of the coils  $D$ , the spot can be made to scan the screen to produce a television frame.

It has been found in practice that more parameters are needed for adjusting the electron gun, especially if one and the same type of tube is to be used for all the different line scans of European television transmissions. We have therefore designed an electron gun with five electrodes and this pentode gun is now being used in nearly all Philips picture tubes. We shall be better able to discuss the possibilities of the pentode gun by first considering some general matters of design and electron optics. Finally, we shall describe in more detail the practical construction of the gun.

### The requirements to be met by the gun and the picture formation

The first requirement is a picture of adequate brightness. As stated, at moderate currents the brilliance with a given screen material is proportional to the power of the beam. This more or less fixes the requisite value of the product  $I_s V_{g_3}$ . Further, the spot on the screen should be so small as to allow the sharpness of the picture to be determined exclusively by the television system, i.e. by the number of lines per frame and the bandwidth. This is the case if the lines of the frame can only just be separately distinguished on the tube screen. The spot becomes smaller as  $V_{g_3}$  increases and  $I_s$  decreases. If the price is to be kept down, both values will be separately fixed by these factors. Moreover, the spot should only be slightly larger in the corners of the screen than in the centre. Since modern tubes are required to be short in order to keep the size of the television cabinet within manageable proportions, the gun itself must be short and larger angles of deflection must be used <sup>1)</sup>.

Now, spot size increases with the deflection angle. This deflection-defocussing has two causes:

- 1) The curvature of the focal plane is greater than that of the picture tube screen. If the middle of the picture is well focussed, the focus at the edges lies behind the screen (i.e. inside the tube).
- 2) An increasing angle of deflection causes increasing astigmatism. The magnitude of this aberration depends upon the design of the deflection coils.

Both defocussing effects are reduced the smaller the diameter of the beam where it passes between the deflection coils. There is therefore a prescribed maximum for the diameter  $d_D$  in fig. 3: the beam must not exceed a certain specified width, and this is largely determined by the degree of prefocussing and the position of the focussing lens. The question now is what means should be adopted to obtain a narrow beam and what is the optimum that can be achieved.

### Electron-optical considerations

#### The Abbe sine condition

The Abbe sine condition states that if a small plane object of area  $S$ , situated in a region of refractive index  $n$  forms an image of area  $S'$  in a region of refractive index  $n'$ , the semi-angles of the cones of the emergent and incident rays being  $u$  and  $u'$  respectively (see fig. 4).

<sup>1)</sup> For a more detailed treatment see J. L. H. Jonker, Philips tech. Rev. 14, 361-367, 1952/53.

$$n^2 S \sin^2 u = n'^2 S' \sin^2 u' \dots (1)$$

A similar postulate applies in electron optics. In the case of electron lenses, the potential  $\Phi$  is substituted for the square of the refractive index, thus:

$$\Phi S \sin^2 u = \Phi' S' \sin^2 u' \dots (1a)$$

(the quantities corresponding to the image are again designated by dashes). The potential is determined with respect to the plane where the electron velocity is zero, i.e. near the plane of the cathode.

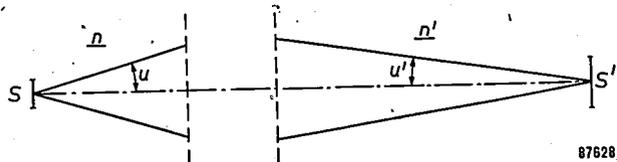


Fig. 4. Diagram to illustrate the Abbe sine condition.

By analogy with optics, we can introduce the quantity "electron-optical brightness"  $B$ , substituting the total beam current  $I_s$  for the luminous flux. We can formulate this brightness as:

$$B = \frac{I_s}{\pi S \sin^2 u}$$

If we compare the brightness  $B'$  in the image space with the brightness  $B$  in the object space, we see from (1a) that:

$$\frac{B'}{\Phi'} = \frac{B}{\Phi}$$

The brightness can therefore be increased by raising the voltage.

In order to be able to apply (1a) to a picture tube, we must know what is acting as the object in the tube and what are its potential and its diameter.

*Characteristics of the cross-over*

The focus of the electron beam is an image (brought about by the focussing and prefocussing lenses) of a point of intersection in the beam; this we shall call the (first) cross-over. It is formed in the triode section of the electron gun, and we shall therefore consider this section separately.

The paths of a number of electrons in the triode section are represented schematically in fig. 5. Each point of the cathode emits electrons in all forward directions, so that each point can be said to emit a separate beam. These beams are bent by the electric field and made to converge. The axis of an individual beam is called the principal ray; at some point along this the rays of the beam converge. This point is the image of the corresponding point on the cathode, and we therefore obtain the image  $k'$  of the cathode  $k$ .

Between  $k$  and  $k'$  the principal rays intersect the axis at the position marked  $C$  in fig. 5. This is the cross-over. We see that the cross-section of the total beam emitted by  $k$  is smallest at this position, being almost equal to the vertical cross-section of a beam originating from a single point of  $k$ . The image of the cross-over appears on the screen, and it is to this image that we apply the Abbe sine condition (see 1a). All quantities referring to the cross-over are suffixed  $c$ , and those referring to the screen image are suffixed  $s$ . Furthermore, the sine is replaced by the angle itself; this is justified by the small beam angles in these tubes (paraxial approximation). The formula now becomes:

$$S_s \Phi_s u_s^2 = S_c \Phi_c u_c^2 \dots (2)$$

Finally, we introduce a quantity  $K$ , defined by

$$\frac{1}{K} = S_c \Phi_c u_c^2 \dots (3)$$

We may therefore write

$$S_s u_s^2 = \frac{1}{\Phi_s K}$$

or, because  $S_s = \frac{1}{4} \pi d_s^2$ :

$$u_s d_s = \sqrt{\frac{4}{\pi \Phi_s K}} \dots (4)$$

As we have seen, the screen potential  $\Phi_s$  is predetermined ( $\Phi_s = V_{g3}$ ). If we consider  $K$  as given, the product  $u_s d_s$  is also determined.

The diameter of the spot  $d_s$  should, as stated, be sufficiently small. On the other hand, this should not give rise to too much deflection defocussing which, for a given size of tube, increases with increasing beam angle  $u_s$ , i.e. is greater the wider the beam. Since, according to (4), the product  $u_s d_s$  is fixed, we must strike a compromise here. Only by increasing the value of  $K$  can we reduce both the

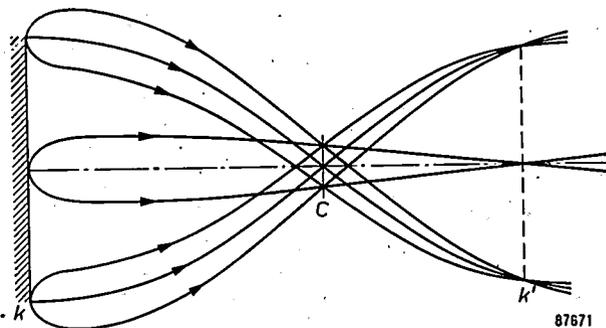


Fig. 5. The electrons from cathode  $k$  form an image of the cathode at  $k'$ . Owing to emission in all directions, each point of  $k$  emits a beam of finite width. The axes of these beams intersect at  $C$ , at which position the total beam is narrowest in diameter. This point is called the cross-over.

diameter of the spot and the deflection defocussing. The actual manner in which the picture is produced, for example by the number of lenses and their focal length, has no effect.

What is important, however, is that equation (4) is valid only for a perfect image. Owing to geometrical errors in the lenses (spherical aberration and axial astigmatism) and to space-charge effects, the product  $u_s d_s$  is always larger than would follow from (4). (This will be dealt with in more detail under the next heading.) Formula (4) cannot therefore be used for quantitative calculations but it does serve as an indication of how the different quantities influence the solution of the problem and as a means of calculating a lower limit of the  $u_s d_s$  product.

The value of  $K$  is solely dependent upon the characteristics of the cross-over (see (3)) and can in principle be derived from calculations of electron paths in the triode section. The calculation is greatly complicated by the fact that space charge effects cannot be neglected. However, by introducing certain idealisations, we can nevertheless make certain statements about  $K$ .

If there is uniform emission over the whole surface of the cathode and if the emitted electrons have a Maxwellian velocity distribution, the maximum current density in the cross-over (by paraxial approximation) will be, as shown by Langmuir<sup>2)</sup>:

$$j_{c \max} = j_0 \left( \frac{e \Phi_c}{kT} + 1 \right) \sin^2 u_c,$$

where  $j_0$  is the current density at the cathode,  $T$  the absolute cathode temperature,  $e$  the charge on the electron and  $k$  the Boltzmann constant. If, as is usual with these tubes,  $e \Phi_c \gg kT$  ( $kT/e \approx 0.1$  V for  $T = 1100^\circ\text{K}$ ) and  $u_c \ll 1$ , we can write:

$$j_{c \max} = \frac{e \Phi_c}{kT} j_0 u_c^2. \quad \dots \quad (5)$$

Assuming uniform current density in cross-over, the total current will be  $I_s = j_c S_c$ , and it therefore follows from (3) and (5) that:

$$K = \frac{e j_0}{kT I_s} = \frac{e}{kT} \frac{1}{S_0}. \quad \dots \quad (6)$$

where  $S_0$  is the effective cathode surface.

However, the current density in the cross-over is not uniform. From fig. 5 we see that the principal rays converge in the middle of the cross-over. The

peripheral parts of the cross-over are formed by electrons emitted with lateral velocities. Starting from this premise, Dosse<sup>3)</sup> calculated that the average current density of the cross-over as well as  $K$  are smaller by a factor of 3 than as indicated by (6).

A second correction is needed to allow for the limitation of the emission by the space charge. Owing to the space charge, the emission from the edge of the cathode is lower than in the middle. Using a simplified theory, Ploke<sup>4)</sup> estimated that the total cathode current is smaller by a factor of 2.5 than it would be if there were uniform emission from the whole cathode surface. Taking  $j'_0$  to be the current density in the middle,

$$I_s = 0.4 S_0 j'_0.$$

Since all points of the cathode contribute to the density of each point in the cross-over (see fig. 5) this density too is reduced by a factor of 2.5.

Owing to both causes,  $K$  is finally found to be:

$$K = \frac{1}{7.5} \frac{j'_0}{I_s} \frac{e}{kT}. \quad \dots \quad (7)$$

The value of  $I_s$  being predetermined (see above) it follows from these formulae that in order to obtain the largest possible  $K$  — and thus the lowest possible  $u_s d_s$  product according to (4) — the current density as the cathode should be high and the cathode temperature as low as practicable.

#### Some practical data compared with theory

To test the theory given in the foregoing we shall now make a simple calculation on the basis of data used in the design of an actual picture tube. Under adverse conditions (bright environment) the average picture-brilliance needed is about 100 nit (i.e. 29 ft L.) For a picture area of about  $0.1 \text{ m}^2$  (e.g. a rectangular tube with a diagonal of 43 cm or 17"), the screen emitting diffuse light, this entails a luminous flux of  $100 \times 0.1 \times \pi = 31$  lumens. The luminous efficiency of a fluorescent screen (without metal backing) can be put at approximately 10 lm/watt, so that the average power of the electron beam must be about 3 watt.

The voltage should be chosen as high as possible, having regard to spot sharpness (see equation 2). The practical limits are set here by safety requirements and considerations of size and cost price. In the tube to be described the accelerating voltage is 14 kV. To obtain a power of 3 watt we therefore need

<sup>3)</sup> J. Dosse, Z. Phys. 115, 530-556, 1940.

<sup>4)</sup> M. Ploke, Z. angew. Phys. 3, 441-449, 1951, and 4, 1-12, 1952.

<sup>2)</sup> D. B. Langmuir, Proc. I. R. E. 25, 977-991, 1937.

an average current of 0.2 mA. The permissible average load on the oxide cathode used is about  $0.5 \text{ A/cm}^2$ . If the cathode temperature is also known, we now have the data required for determining  $K$ . The temperature of the oxide cathode being about  $1100^\circ\text{K}$ , the value of  $K$ , according to (7), is:

$$K = 3.3 \times 10^7 \text{ V}^{-1} \text{ m}^{-2}.$$

Assuming further that the beam angle at the screen is restricted to  $2 \times 10^{-3}$  rad in order to avoid excessive deflection-defocussing, we can calculate the theoretical diameter  $d_s$  of the spot on the screen as 0.8 mm. In practice, however, diameters are found that are 1.5 times as large.

We can try to ascertain what causes this widening of the spot. Our electron-optical considerations lose their validity if the electron paths can no longer be considered as paraxial and if space charge phenomena occur. In the first case, spherical aberration can become serious in the triode section and in any prefocussing lens that may be provided. Spherical aberration in the final lens may reasonably be neglected for such narrow beams as these. From Floke's<sup>4)</sup> calculations it appears that a 25% increase in beam diameter can occur in the triode section as a result of spherical aberration. The spot on the screen is therefore enlarged by the same amount.

The inherent space charge of the beam also gives rise to some widening, which is difficult to calculate exactly. An approximative method by Wendt<sup>5)</sup> gives a widening of 15% for the example under discussion.

After accounting for both these errors of image formations, the agreement between theory and experiment is quite satisfactory.

We shall now devote our attention in particular to the electron gun, first entering rather more closely into the question of prefocussing.

#### Closer consideration of prefocussing<sup>6)</sup>

As stated, the image quality of a picture tube is determined to a large extent by the width of the beam in the deflection coils: the less the width the less the deflection defocussing (see the article quoted in 1). The amount of deflection defocussing for a given beam diameter  $d_D$  is determined by the radius of curvature of the screen and by the quality of the deflection coils. Fig. 6 shows the focussed size of spot  $d_s$  as a function of the distance  $l$  from the middle of the screen, for tubes with a screen

diagonal of 43 cm (17") and two widths of beam, narrow and wide respectively. Apart from the gun, both tubes are identical. Curve *a* refers to the narrow beam, curve *b* to the wide. We see that the tube with

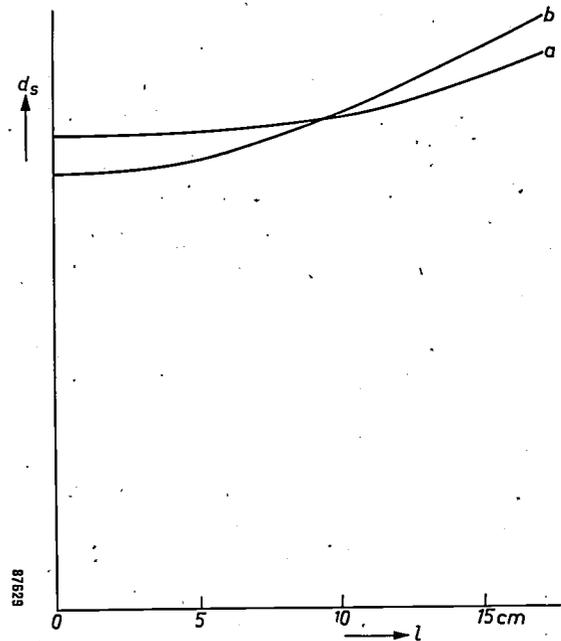


Fig. 6. The diameter  $d_s$  of the focussed spot as a function of its distance  $l$  from the middle of the screen; *a* for a narrow beam, *b* for a wide beam.

the wide beam is better in the middle than the tube with the narrow beam. This agrees with the paraxial theory, according to which  $d_s$  in the middle of the screen is inversely proportional to  $u_s$  — the parameters of the cross-over and the anode potential  $\Phi_s$  being constant. However, the narrow beam (curve *a*) gives a more sharply focussed picture at the edge of the screen. In addition to deflection defocussing, there is usually some "modulation defocussing" to be reckoned with. This may be accounted for as follows. If, after having focussed the electron beam at a specific current, we proceed to increase the current (in practice by varying the voltage on the control grid) two effects are produced. In the first place the cross-section of the cross-over (and thus of the "object") becomes larger in the triode lens as a result of spherical aberration and space charge phenomena, and therefore the spot on the screen also becomes larger. In the second place the reduced power of the triode lens with a less negative grid causes the cross-over to shift towards the screen, so that the position of the object is changed and the beam is no longer focussed on the screen. Modulation defocussing is greater the narrower we make the beam. This can be seen in fig. 7 where the size of the spot in the middle of the screen is plotted as a function of the beam current  $I_s$  (drawn lines), the tube

<sup>5)</sup> G. Wendt, Ann. Physik 2, 256-264, 1948.

<sup>6)</sup> An important contribution to the development about to be described was made by W. F. Niklas, now with Columbia Broadcasting System.

again being the 43 cm type and the focussing magnet being set for one specific beam current, in this case 100  $\mu$ A. The broken lines indicate the size of spot when the beam is focussed for each value of the current, which is, of course, not the case in practice.

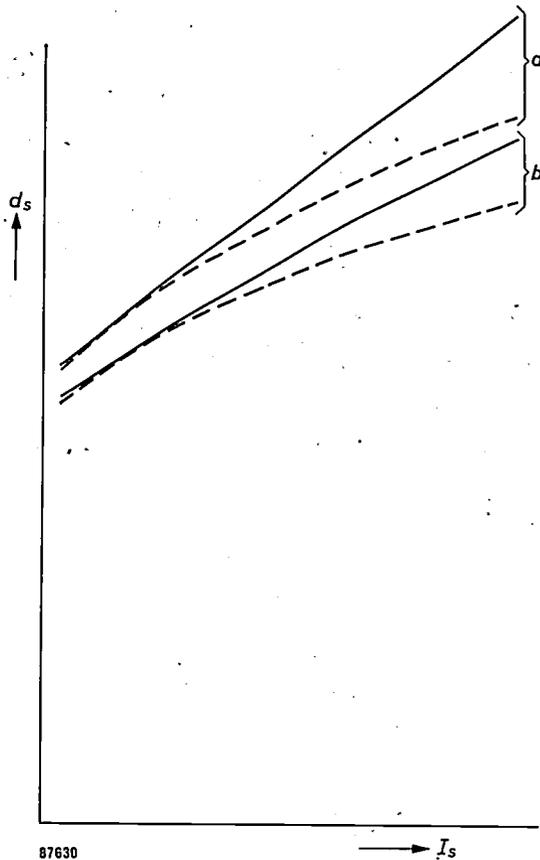


Fig. 7. The spot diameter  $d_s$  at the middle of the screen as a function of the beam current  $I_s$ ;  $a$  for a narrow beam,  $b$  for a wide beam.

Curves  $a$  and  $b$  refer to a narrow and wide beam respectively. The average slope of these curves is thus a measure of the modulation defocussing, and

we see that the slope is somewhat flatter for the wide than for the narrow beam. The explanation of this phenomenon is too complicated to be included in this paper.

It will be shown below that the use of a pentode gun allows us in a given tube to vary within certain limits the compromise between the definition at the edge and that in the middle of the screen.

### The position of the prefocussing lens

We have shown that, according to the paraxial theory, it is immaterial where the prefocussing takes place. However, there are three reasons for positioning the prefocussing lens as close as possible to the triode section:

- 1) The closer the prefocussing lens to the triode section, the greater will the power of the lens have to be in order to obtain the same width of beam. It is known that spherical aberration in electron lenses generally decreases according as the focal length is reduced.
- 2) It is advisable to limit as much possible the area in which the electrons travel at a lower velocity and in which, therefore, the space charge causes greater divergence.
- 3) Owing to constructional restrictions in the diameters of the lens electrodes, it is advisable to keep the beam width in the prefocussing lens as small as possible in order to minimize aberrations of the image.

### The form of the prefocussing lens

Some possible forms of prefocussing lens are shown in fig. 8a, b and c. The version generally adopted in television tubes until fairly recently is represented in fig. 8a. There are two parameters

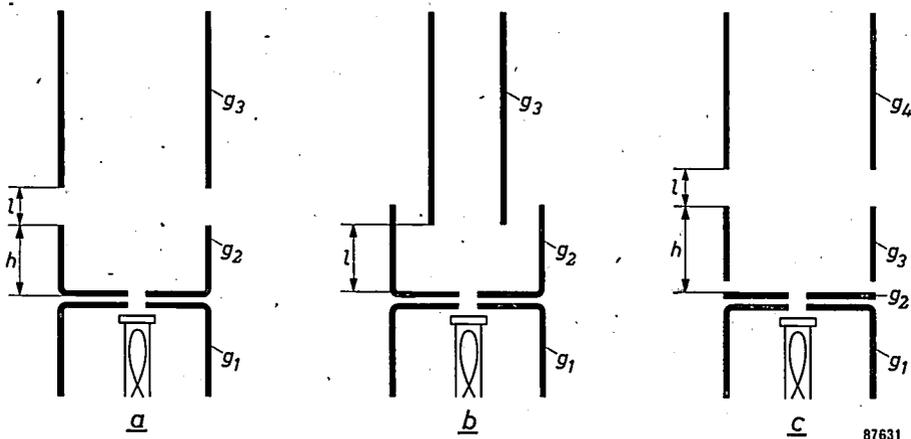


Fig. 8. Three possible versions of a gun with prefocussing lens. a) Tetrode gun, in which  $g_2$  and  $g_3$  have the same diameter, b) the same, but  $g_2$  and  $g_3$  have different diameters, c) pentode gun.

that are mainly responsible for prefocussing the beam, viz. the distance  $l$  between the first anode  $g_2$  and the final anode  $g_3$ , and the height  $h$  of the first anode.

The power of the lens is also affected by the diameter of the first anode; a smaller diameter gives a more powerful lens. In practice, however, the choice of diameter is restricted, outwardly by the size of the neck and inwardly by the risk of spherical aberration.

By changing the distances  $h$  and  $l$ , the power of the lens and hence the focal length are changed. Both  $h$  and  $l$ , when increased, cause greater prefocussing, but an increase in  $h$  has more effect than the same increase in  $l$ . To keep the lens short,  $l$  is made as small as possible without running the risk of field emission — the potential difference between the electrodes  $g_2$  and  $g_3$  can be as much as 20 kV — while the value of  $h$  to give the requisite prefocussing is determined. A useful measure of the prefocussing, and one that can easily be verified experimentally, is the diameter of the spot on the screen,  $d_v$ , which is obtained without bringing the focussing lens  $f$  into action (see fig. 3). With a given type of tube, the diameter  $d_D$  of the focussed beam in the deflection coils is reasonably proportional to  $d_v$ . The diameter  $d_D$  has an important bearing on the amount of deflection and modulation defocussing, as described. Fig. 9a shows  $d_v$  as a function of  $l$ , for various values of  $h$ .

A second possible prefocussing lens is represented in fig. 8b; here the final anode is smaller in diameter than the first anode and is enclosed by the latter. The power of the lens is now determined by the distance  $l$  and the ratio of the diameters of both anodes. Although this solution has been put into practical use, it has the drawback that the distance  $l$  is difficult to adjust during assembly owing to the first anode being entirely enclosed by the final anode. Fig. 9b shows the spot diameter  $d_v$  as a function of the distance  $l$ , for the case in which the ratio of the diameters is 1 : 2.

A third possibility is given in fig. 8c. This principle, which is now being applied in the design of almost all Philips picture tubes, entails the introduction of a fifth electrode. Suppose that in fig. 8a we cut through the first anode just above the base and then place the separate cylinder thus obtained a small distance away from  $g_2$ . If we now bring the annulus  $g_2$  and the cylinder  $g_3$  to the same potential as  $g_2$  in fig. 8a, there will be practically no change in the action of the lens, when the distance between  $g_2$  and  $g_3$  is very small. However, if the potential on the cylinder is made to differ from that on  $g_2$ , the power of the

lens will certainly change; the prefocussing lens will be more powerful when  $V_{g_3} < V_{g_2}$  and weaker when  $V_{g_3} > V_{g_2}$ . If  $V_{g_2}$  remains unchanged, the modulation characteristic of the tube is not affected. We now have an electron gun with variable prefocussing, so that by regulating the voltage on  $g_3$  we can achieve the best compromise between spot size and deflection defocussing.

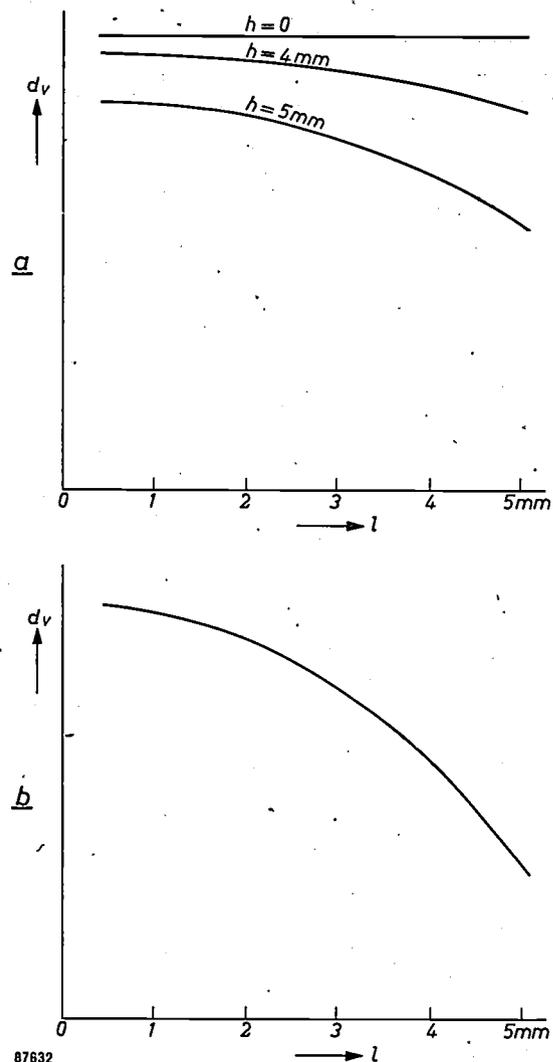


Fig. 9. a) The diameter  $d_v$  of the unfocussed spot as a function of the distance  $l$  between  $g_2$  and  $g_3$  for a gun as in fig. 8a. The parameter  $h$  is the length of  $g_2$ . b)  $d_v$  as a function of  $l$  for a gun as in fig. 8b.

### Practical version of the pentode gun

During the design of the new gun incorporating the prefocussing lens as in fig. 8c, the voltage on  $g_2$  was found to have a considerable influence on the power of the lens. This had to be avoided, since it would leave the designer less freedom in his choice of  $V_{g_2}$ . The method adopted was to provide the cylinder  $g_3$  with a base, as shown in fig. 10a, as a result of which the field of  $g_2$  is prevented from

penetrating so far into the space between  $g_3$  and  $g_4$ . Fig. 10a illustrates the layout of the electrodes in the gun finally constructed. Electrode  $g_3$  is slightly rounded and fitted with a plate to counteract field emission to the high-potential electrode  $g_4$ . All electrodes are mounted on ceramic rods, one of which can be seen in the figure <sup>7)</sup>. Electrode  $g_4$  is bent off-axis to provide, in conjunction with a permanent magnet, an ion trap <sup>8)</sup>. The lower end of this electrode is terminated by an annular cap which determines the correct separation from the other electrodes. Springs are fixed to the top end for

obtained without difficulty from the power pack. The voltages used in practice are therefore between  $-50$  and  $+350$  V. Since  $g_3$  draws no current, its potential can readily be adjusted with a potentiometer. In this way, the width of the beam can be varied by a factor of 1.5, which is more than enough for the beam widths needed in practice.

Owing to the differences in the line systems and bandwidths used in European countries it is desirable to be able to vary the size of the spot on the television screen. For the English system of 405 lines the spot on the screen can be somewhat larger than

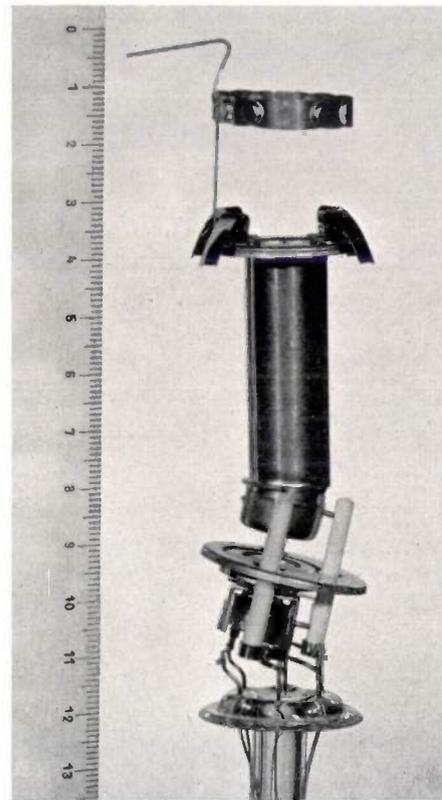
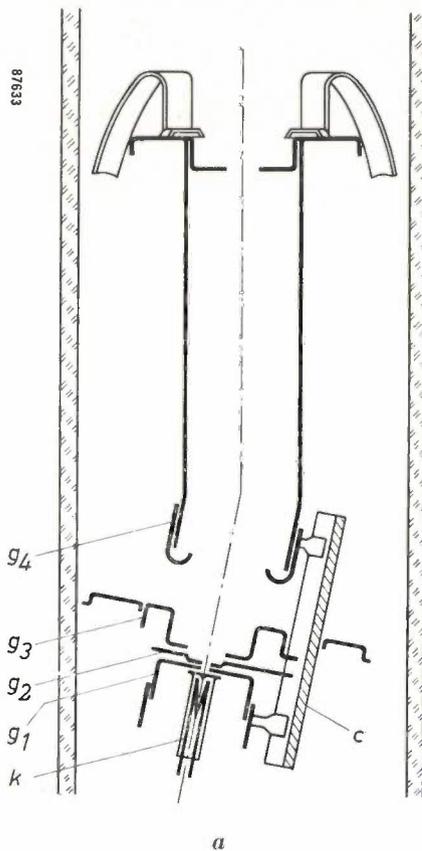


Fig. 10. a) Structure of a pentode gun. The gun is mounted on three ceramic rods *c*. Electrode  $g_4$  is centred with respect to the glass envelope by springs at the top. The bend in this electrode, in conjunction with a permanent magnet, serves as an ion trap. b) Photograph of the finished gun.

the purpose of centring the electrode in the glass envelope. A photograph of the gun is shown in fig. 10b. The ring visible above electrode  $g_4$  contains a getter.

The voltage on  $g_2$  can now be varied within fairly wide limits without seriously affecting the power of the prefocussing lens. The electrode  $g_3$  must be so designed that the voltage on this electrode can be

for the Gerber system of 625 lines, and an even smaller spot is needed for the French system of 819 lines. In a television receiver intended for the English system, the prefocussing electrode  $g_3$  should, as a rule, be given cathode potential ( $V_{g_3} = 0$  V) which results in a larger spot. The strong prefocussing further produces a narrow beam, the advantage of which is a spot of very uniform size all over the screen, even at the edges (very little deflection defocussing). This allows us to simplify the deflection coils to some extent, which slightly increases the

<sup>7)</sup> This construction is described by J. de Gier and A. P. van Rooy in Philips tech. Rev. 9, 180-184, 1947/48.

<sup>8)</sup> See W. F. Niklas, Philips tech. Rev. 15, 258-262, 1953/54.

deflection defocussing but does not impair the quality of the picture. The designer may then devote rather more attention to obtaining a good rectangular form of raster, a requirement opposed to that of small deflection defocussing. This setting does, however, cause a slight increase in modulation defocussing. With the beam currents concerned, the spot is not noticeably enlarged thereby, but it is a reason for not decreasing  $V_{g3}$  still further.

For the 819 line system the procedure is just the opposite. Electrode  $g_3$  is given a potential of about 300 V (e.g. by connecting  $g_3$  with  $g_2$ ). The spot in the middle of the screen is then as small as possible and there is also very little modulation defocussing. Owing to the beam being wider, however, the designer has to pay closer attention to coil design in order to avoid excessive deflection defocussing.

It is clear that the pentode gun allows the set designer to vary the properties of the gun to suit his design. The gun is little more complicated than its predecessors, which means that the additional facilities it offers need hardly affect the price of the set.

Fig. 11 shows the size of the focussed spot  $d_s$  (broken lines) and of the unfocussed spot  $d_v$  (drawn lines, on a  $10\times$  larger scale) as a function of the prefocussing potential  $V_{g3}$ , at a beam current of  $100\ \mu\text{A}$ , for the three values of  $V_{g2}$ . As we have said,  $d_v$  is a good measure of the beam width in the deflection coils. The figure confirms that with increasing  $V_{g3}$  the beam becomes wider and the spot smaller. The curves given were plotted from data obtained from a type MW 43-64 tube; analogous curves apply to other types of tube, e.g. MW 36-44 and MW 53-20.

To conclude, some other interesting aspects of the pentode gun deserve mention, even though their value may be only theoretical. As explained in the foregoing, it would be desirable, having regard to modulation defocussing, to be able to focus the electron beam for each value of beam current.

It is obvious that, with a permanent focussing magnet as frequently used at the present time, such additional adjustment is not practicable. When using the pentode gun, however, the possibility exists of making the appropriate correction by means of an auxiliary voltage on  $g_3$  derived from the modulation signal.

A similar means of reducing deflection defocus-

ing would be to modulate  $g_3$  with a voltage derived from the deflection currents — dynamic focussing. Naturally, this would only reduce the errors due to field curvature; it would not correct astigmatism. Both improvements require an extra circuit, which will probably be found too intricate for general use.

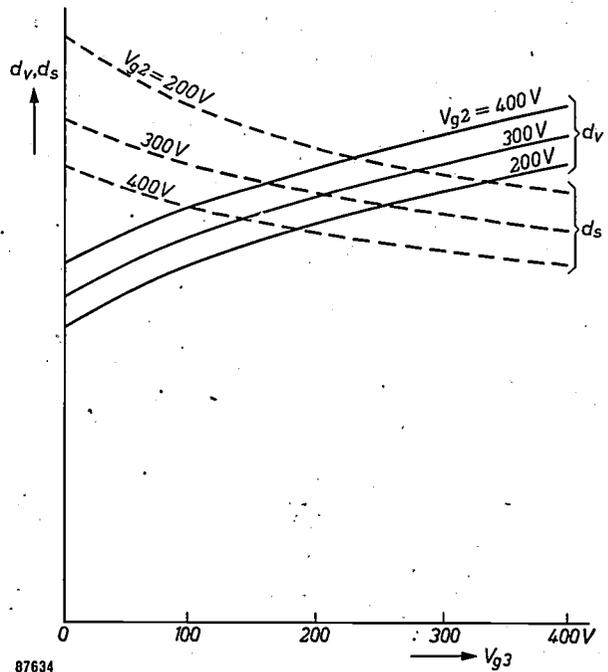


Fig. 11. The diameter  $d_v$  of the unfocussed spot (full lines) and  $d_s$  of the focussed spot (broken lines) as a function of  $V_{g3}$ , for various values of  $V_{g2}$  ( $d_s$  on a scale  $10\times$  larger than  $d_v$ ).

**Summary.** When designing a television picture tube it is necessary to make the spot on the screen so small that the definition of the picture is determined by the television raster alone. This imposes certain demands upon the electron gun, which forms the beam. The beam is required to be as narrow as possible, because deflection defocussing increases with the diameter of the beam in the deflection coils. Consideration of the electron optics reveals the principal possibilities. Only the product of beam angle and spot size occurs in the equations and this product becomes smaller with increasing screen potential and current density at the cathode. If spherical aberration and space charge phenomena are taken into account, the values obtained in practice are in fairly good agreement with the theory.

Closer consideration is then given to the prefocussing lens. This should be positioned as near as practicable to the cathode. Examination of various possible forms of prefocussing lenses has led to the development of a pentode gun, which is now being used in almost all Philips picture tubes. Varying the voltage on the third electrode changes the width of the beam and the size of the spot on the screen, thus enabling the optimum size of spot to be chosen to suit the line system in use. It also allows a very narrow beam to be employed with a fairly coarse line system, which means that the deflection coils can be simplified without causing excessive deflection defocussing. Finally, the pentode gun makes it possible in principle to apply partial correction of deflection and modulation defocussing.

## AMPLITUDE MODULATION OF CENTIMETRE WAVES BY MEANS OF FERROXCUBE

by H. G. BELJERS.

621.376.2:621.396.029.64:621.318.134

*The Ferroxcube discussed in this article is mainly used as magnetic core material for frequencies up to about 100 Mc/s. At higher frequencies, phenomena occur which prevent its use as core material but which have led to a further, interesting application. This is one of those cases not infrequently occurring in technology where a particular discovery or a certain product is found to be applicable in a field different from that for which it was originally intended.*

Radio waves with a wavelength of less than about 30 cm have acquired considerable importance over the last ten years or so for experimental work in widely different scientific fields. Some examples are: microwave spectroscopy, gyromagnetic resonance phenomena and radio-astronomy. The most important industrial applications are radar and beamed communication links. For distances of 10 km or more absorption in the atmosphere precludes the use of wavelengths of less than 3 cm.

In radio communication, the transmission of information is effected by modulating a carrier wave and in principle we may use either frequency modulation (FM) or amplitude modulation (AM). Amplitude modulation may be effected by varying the voltage on one of the electrodes of a transmitting tube but it is generally accompanied by an interfering frequency modulation. Consequently it is preferable to bring about amplitude modulation in another manner e.g. by absorption. In this method the modulation is effected outside the transmitting tube; a varying fraction of the HF power is absorbed in the output circuit.

Study of the magnetic properties of ferrites has shown that in an H.F. magnetic field these materials are subject to losses dependent upon the frequency and the type of ferrite and also that these can be influenced by an external magnetic biasing field<sup>1)</sup>. This work, originally extending to frequencies of a few hundred Mc/s, has since been extended to much higher frequencies and it has been shown that the magnetic losses still occur up to about 10,000 Mc/s. From observations by W. J. van de Lindt in the Philips Laboratories at Eindhoven it has been found that, with a number of ferrites, weak external magnetic fields can produce large variations in these losses; hence by low frequency variation of the external field we should, in principle, be

able to bring about an absorption modulation<sup>2)</sup>.

To achieve a modulation of this type a small ferroxcube rod may be fitted into a cylindrical cavity resonator as shown in fig. 1a at a point where the alternating magnetic field is strong. Fixed round the rod are coils by means of which a modulating magnetic field can be excited in the axial direction. The cavity resonator is coupled to a wave-guide through which the H.F. energy to be modulated flows. If the cavity resonator is tuned to the frequency of the wave in the guide (by adjusting the

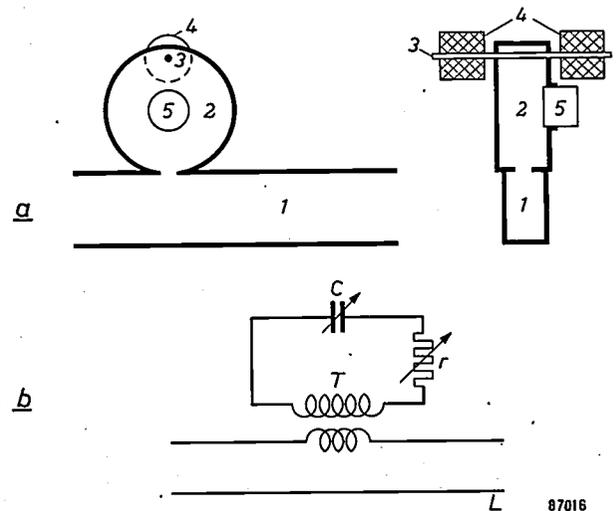


Fig. 1. Arrangement for "absorption modulation". a) the ferrite rod 3, fixed in the cavity resonator 2 which can be tuned by the plunger 5, can be magnetized by the coils 4, thereby causing variations in the losses in the cavity resonator. Owing to the coupling with the wave-guide 1, the power transmitted in the wave-guide varies.

b) Equivalent circuit of the arrangement. The cavity resonator, represented as a series resonant circuit coupled to a line  $L$ , is tuned by the variable capacitor  $C$ . The coupling between wave-guide and cavity is represented by the transformation ratio of the transformer  $T$ . The variable losses are represented by the variable resistor  $r$ . When  $r$  varies, the  $Q$  of the circuit changes, thereby altering the impedance in the line  $L$  and thus the amount of reflection that occurs.

<sup>1)</sup> H. G. Beljers and J. L. Snoek, gyromagnetic phenomena occurring with ferrites, Philips tech. Rev. 11, 313-322, 1949/50.

<sup>2)</sup> H. G. Beljers, W. J. van de Lindt and J. J. Went. A new point of view on magnetic losses in anisotropic bars of ferrite at ultra high frequencies, J. appl. Physics, 22, 1506, 1951.

plunger 5) the energy is partially reflected at this point: the higher the  $Q$  of the cavity resonator, the greater the amount reflected. The equivalent circuit fig. 1b, illustrates the arrangement further. By varying the current through the coils, the  $Q$  of the cavity resonator can be altered since the changing magnetic field causes variations in the losses of the ferroxcube. This, in turn, varies the amount of reflection and hence the amplitude of the wave passing through the wave-guide is modulated. Although the modulation cannot be attributed to direct absorption of the H.F. power to be transmitted but to a reflection controlled by absorption of a fraction of the power, we shall continue to refer to it here as absorption modulation.

Ferroxcube IVA and ferroxcube IVB have been used with success in this manner for a wavelength of 3.2 cms (9,300 Mc/s). In the following we shall show that the saturation magnetization of the ferroxcube determines the highest carrier frequency at which absorption modulation is possible. So far, no types of ferroxcube have been found possessing a sufficiently high saturation to permit modulation at carrier frequencies higher than about 12,000 Mc/s.

Using the experimental arrangement for 9,300 Mc/s mentioned above, a reasonable modulation was attained up to a modulation frequency of 1.3 Mc/s. At higher modulation frequencies it proves difficult to excite the desired low frequency field in the ferroxcube because the capacitance between the windings of the coils becomes considerable and leads to a weakening in the field.

Although, therefore, this modulation method cannot be considered for bandwidths larger than a few Mc/s it nevertheless affords interesting possibilities for using ferroxcube at very high frequencies<sup>3)</sup>.

The phenomena that arise will be discussed in more detail below.

#### Magnetic losses in ferrites

The magnetic losses of ferrites in an H.F. field may be regarded as being a result of gyromagnetic resonance.

Consider a body with a certain spontaneous, magnetic moment. Such bodies do exist, viz. the so-called Weiss domains in a ferromagnetic material (regions of dimensions of  $0.1 \mu$  to  $10 \mu$ ) where all electron spins are roughly in the same direction. If the body is placed in a constant magnetic field  $H$  making a certain angle with the magnetic moment,

then, as explained in<sup>1)</sup> the vector of the magnetic moment precesses about the field direction with an angular frequency  $\omega_p$  where

$$\omega_p = \gamma H. \quad (1)$$

The constant  $\gamma = 2.2 \times 10^5$  m/Asec is termed the gyromagnetic ratio. This relation holds so long as the precessing magnetic moment itself has no effect on the effective field  $H$  (i.e. if no demagnetization occurs). This is the case with a spherical body. If we now apply a small HF magnetic field perpendicular to  $H$ , the angle of precession increases but for it to do so, energy has to be supplied from the HF field to the spin system: the magnetic moment is, as it were, bound with a certain coupling to the direction of the magnetic field  $H$ . If the alternating field has exactly the angular frequency  $\omega_p$  of the precession, resonance occurs and the magnetic vector precesses at a greater angle with respect to the field direction, which corresponds to the least amount of energy. The precession energy absorbed when an alternating field is applied, is, however, partially transferred to the crystal lattice by interaction with it and this results in an increase in the temperature of the material: a part of the H.F. power is converted into heat and losses occur in the H.F. field.

Closer investigation shows that in a ferromagnetic material resonance losses occur even without an external constant field. The magnetization in each Weiss domain has a preferential direction, namely that in which the lattice energy is minimum. The magnetization is bound to this state of equilibrium with a certain coupling, which causes it to describe a precessional movement with an angular velocity proportional to the strength of this coupling. Denoting this precession frequency  $\omega_p^{(A)}$ , we can introduce an equivalent field  $H_A$  which for small deflections would give rise to the same coupling, so that we obtain

$$\omega_p^{(A)} = \gamma H_A. \quad (2)$$

If an alternating external magnetic field having the same (or practically the same) angular frequency is now applied, the component of that field perpendicular to the equivalent field  $H_A$  can cause resonance just as in the case where a constant external magnetic field  $H$  is applied. In this case, too, energy is constantly being supplied from the H.F. field to the spin system, thereby giving rise to the development of heat.

The greater the forces due to the crystal structure (so-called anisotropic forces), the more difficult it is to direct the spins by an external magnetic field; the crystal forces are roughly inversely proportional to

<sup>3)</sup> An instance of this was described a short while ago in this review: J. Cayzac, Automatic frequency stabilization for a beam transmitter working on centimetric waves, Philips tech. Rev. 17, 334-339, 1955/56.

the susceptibility. With larger anisotropic forces we have a higher precession frequency; this implies that losses are noticeable only at higher frequencies. In ferrites, crystal anisotropy gives rise to losses which have their maximum value at frequencies between 2 and 200 Mc/s, according to <sup>1)</sup> and experiments confirm that the losses start to appear at higher frequencies at the susceptibility becomes smaller <sup>4)</sup>.

Magnetic losses in ferrites are experimentally found up to about 10,000 Mc/s. This can be understood as follows. In each Weiss domain exists the saturation magnetization  $M$ . When no external magnetic field is present these magnetizations have a different direction from domain to domain, so that viewed as a whole, the material has no magnetic moment. Because the Weiss domains — each separately completely magnetized — vary a great deal in shape, demagnetizing effects will appear, differing in intensity from domain to domain, which can cause the coupling of the magnetic moment with the lattice (and consequently the gyromagnetic resonance frequency) to increase locally by considerable amounts (shape anisotropy).

Consider as an example a magnetized ellipsoid. The demagnetizing effect here is as small as possible provided the magnetic moment  $M$  is in the direction of the major axis (state of least energy). The axis we term the  $z$ -axis. The magnetic moment is bound to this axis with a certain coupling and therefore here too a precession with a particular frequency  $\omega_p'$  can appear. The magnitude of  $\omega_p'$  depends, furthermore, upon the shape of the ellipsoid and it can be shown that the following relation applies:

$$\omega_p' = \gamma \sqrt{[H_A + (N_x - N_z)M][H_A + (N_y - N_z)M]} \quad (3)$$

where  $H_A$  is the equivalent crystal anisotropic field (likewise assumed to be in the  $z$  direction);  $-N_x M$  is the demagnetizing field produced in the ellipsoid if the magnetization is in the  $x$  direction, etc. In general we have,

$$N_x + N_y + N_z = 1.$$

For a sphere  $N_x = N_y = N_z = \frac{1}{3}$ ; we see that  $\omega_p' = \gamma H_A$ ; there is then no preferred direction and the precession frequency is the same as in a field  $H_A$ . With an ellipsoid  $\omega_p'$  is generally higher. We find the highest frequency with an ellipsoid of revolution greatly elongated in the  $z$  direction (needle). For this we have  $N_x = N_y = \frac{1}{2}$  and  $N_z = 0$ , so that here

$$\omega_p' = \gamma(H_A + \frac{1}{2}M) \quad (4)$$

Each Weiss domain in a ferrite shows a resonance curve and corresponding energy absorption round about a frequency of  $\omega_p'$  which depends on the shape of the domain and can be approximately derived from equation (3) which holds for the idealized ellipsoidal shapes. By summing the losses of all domains at one frequency we obtain the losses of the material as a whole at that frequency. Since the needle-shaped domains are probably not very numerous, the losses will decrease with the frequency but will nevertheless exist up to the frequency given by (4). It can be seen from the fact that  $M$  is of the order of  $32 \times 10^4$  A/m (4,000 oersted) whilst the equivalent field  $H_A$  amounts at most to  $1.6 \times 10^4$  A/m (200 oersted) that this frequency is much higher than the resonance frequency corresponding to  $H_A$ . We find, moreover, from a closer analysis that even (4) still does not give the highest possible frequency, but that the precessional magnetisations of lamina-shaped Weiss domains occurring in particular superimposed strata, each with opposing magnetisation, influence each other <sup>5)</sup> and this causes the maximum frequency  $\omega_{p \max}$  at which losses can occur to increase to

$$\omega_{p \max} = \gamma(H_A + M) \quad (5)$$

This frequency is practically twice as high as that derived from (4) and amounts to 4,000—12,000 Mc/s for the various kinds of ferroxcube.

Consider now the influence of an external constant magnetic field  $H$ . We have already seen that with a sphere a resonance frequency  $\gamma H$  occurs. What happens, however, with a thin rod such as is used in the arrangement described for absorption modulation? If we apply here a constant field  $H$  parallel to the axis of the rod, even a small value of this field, e.g. 4,000 A/m (50 oersted) is sufficient to completely saturate the ferrite so that everywhere the magnetization  $M$  is directed along the axis. If we disregard the crystal anisotropy field  $H_A$  that varies locally in magnitude and direction, the whole rod behaves as one large needle-shaped Weiss domain with, according to (4), a resonance in the neighbourhood of the frequency

$$\omega_p^{(H)} = \gamma \cdot \frac{1}{2}M \quad (6)$$

(The crystal anisotropy field  $H_A$ , and in general the external magnetic field  $H$  too, are both small with respect to  $\frac{1}{2}M$ ). The energy absorption of all parts of the rod is now concentrated in this frequency range so that the losses are much higher than in the case where there is no external magnetic field.

<sup>4)</sup> See J. L. Snoek, *Physica* 14, 207-217, 1948.

<sup>5)</sup> D. Polder and J. Smit, *Resonance phenomena in ferrites*, *Rev. mod. Phys.* 25, 89-90, 1953.

**Testing the theory; possible methods of modulation**

The magnetic characteristics of ferrites are usually determined by measuring the complex relative permeability:

$$\mu_r = \frac{B}{\mu_0 H} = \mu' - j\mu'',$$

where the quantity  $\mu''$  is a measure of the losses.

Two curves  $p$  and  $q$  are plotted in *fig. 2*, which show measurements of  $\mu''$  for Ferroxcube IV in the frequency range from 0.2 to 24,000 Mc/s. The external biasing magnetic field is zero for curve  $p$ ; losses

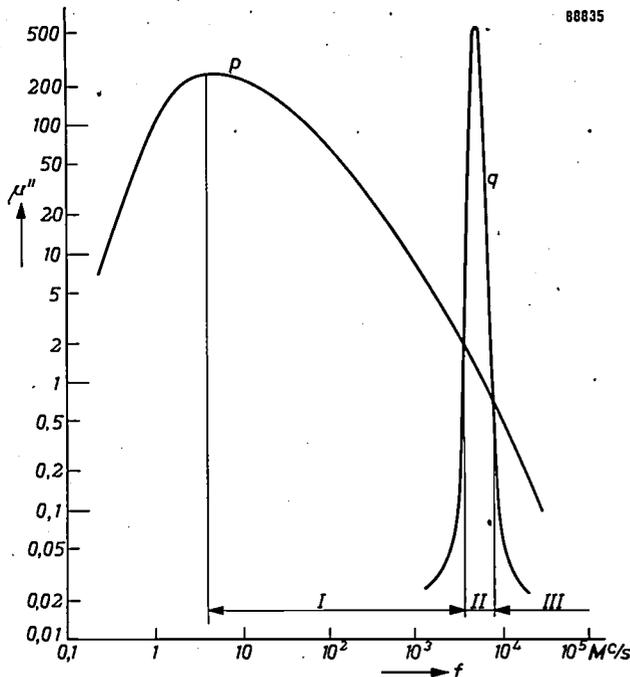


Fig. 2. The losses of an H.F. field in ferroxcube IVA expressed by the imaginary part  $\mu''$  of the relative permeability as a function of the frequency:  $p$  without a magnetic biasing field,  $q$  with a saturating magnetic field.

start here at approximately 0.5 Mc/s and have become relatively small again at 300 Mc/s but nevertheless, still exist up to frequencies of about 9,300 Mc/s. With curve  $q$ , the biasing field is 4000 A/m so that the ferrite rod is saturated. These measurements confirm the theory at least qualitatively. At saturation, the resonance frequency ( $\omega_p^{(H)}$ ) for this ferrite amounts to roughly 5,000 Mc/s. In a demagnetized state, the losses are very small at frequencies higher than about 10,000 Mc/s.

On the basis of these loss curves, it is easy to understand how we can achieve absorption modulation by varying an external magnetic field. In each period of the modulating magnetic field, the loss curve alternates between the form  $p$  and the form  $q$  so that at a fixed carrier frequency the power absorbed varies periodically.

To consider this further, the frequency scale is divided into the regions *I*, *II* and *III* using the abscissae of the intersections of  $p$  and  $q$  as boundaries. Frequency range *I* includes the frequencies from about 100 Mc/s (the modulation method under consideration here is of no interest for lower carrier frequencies) to approximately 3,000 Mc/s. In practice, absorption modulation with ferrites is unlikely to be used in this frequency region since conditions are such that the resonance frequency (peak of loss curve) even at partial saturation by the external magnetic field becomes equal to the carrier frequency. With an increasing magnetic field the losses would first increase and then decrease, thereby giving rise to an undesirable modulation characteristic.

Absorption modulation is also not particularly suitable for region *II* since the losses do not drop to a low value for any value of the magnetic field when the latter remains small with respect to  $M$ . Absorption modulation is best suited to region *III*, viz. the highest frequencies, since with an increasing magnetic field the losses decrease steadily and drop to a low value with a saturating field.

In this connection, we must still bear in mind that  $M$  (and consequently  $\omega_p^{(H)}$ ) depends upon the temperature. If we wish the arrangement to maintain approximately the same modulation characteristic (in range *III*) in a broad temperature range, we must therefore remain sufficiently far away from the resonance peak. Because many kinds of ferroxcube exist having a different saturation

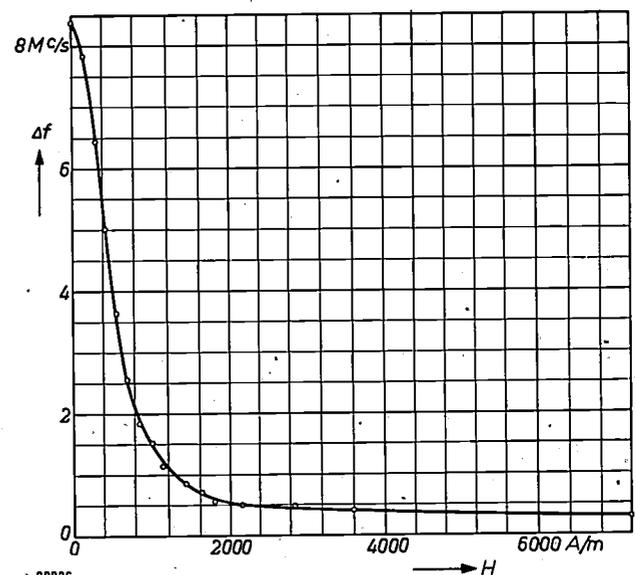


Fig. 3. The losses in an Mn-Mg ferrite as a function of the biasing magnetic field  $H$  at 9,300 Mc/s. The half-value width  $\Delta f$  of the resonance curve of the cavity resonator in which the ferrite is placed is plotted as the ordinate. This  $\Delta f$  is proportional to the losses and hence to  $\mu''$ .

magnetization it is, however, possible to find the right ferroxcube for this modulation method for each carrier frequency between 2,000—12,000 Mc/s. In fig. 3 the losses for a rod of Mn-Mg ferrite at 9,300 Mc/s are plotted as a function of  $H$ ; this is therefore the modulation characteristic for that carrier frequency.

A further peculiarity may be noted here. It is a well-known fact that in the case of magnetic resonance the susceptibility is negative at frequencies above the resonance frequency (the magnetisation is then in anti-phase to the external field). When the resonances of all Weiss domains are superimposed on each other, the permeability  $\mu'$  of the whole ferrite rod at certain frequencies therefore become smaller than unity. This is, in fact, the case close to the upper limit of the loss range, as can be seen a fig. 4 where the real part  $\mu'$  of the permeability for ferroxcube IVA is plotted as a function of the frequency over a wide frequency range. We observe that  $\mu'$  decreases rapidly whenever losses occur (cf. curve  $p$  in fig. 2), and before  $\mu'$  approaches unity at very high frequency there is a range where  $\mu' < 1$ . (In order to delineate the curves clearly in each range, the scale of the ordinate is linear between 0 and 1 and logarithmic for values  $> 1$ ).

Finally, we may give a quantitative experimental confirmation of formula (5) which shows the relation between the maximum frequency at which losses occur in a demagnetized state and the magnetization  $M$  of the Weiss domains (saturation magnetization). Working with large frequency variations in the centimetre wave region is inconvenient; hence, instead of varying  $\omega$ , we may vary the saturation

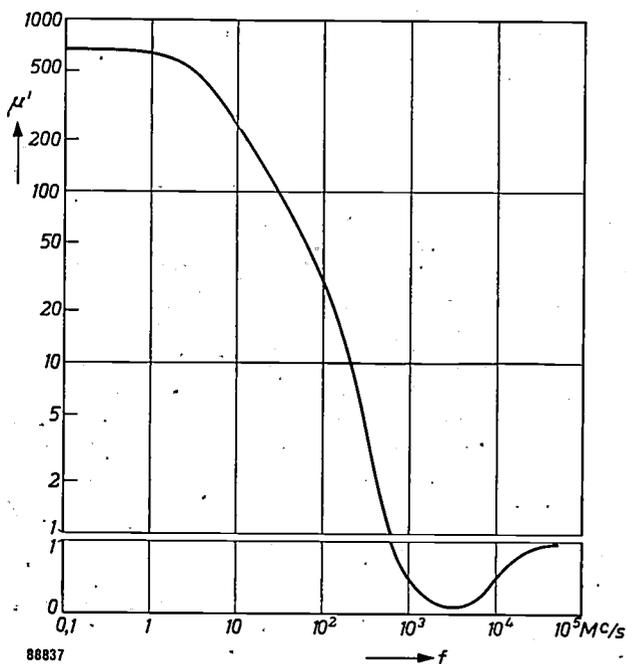


Fig. 4. The real part  $\mu'$  of the relative permeability of ferroxcube IVA as a function of the frequency. Values below unity are plotted on a linear scale; above unity the ordinate is logarithmic.

magnetization  $M$ , keeping the frequency constant. As already mentioned in the foregoing, this saturation magnetization depends on the temperature, as found from other measurements: the magnetization decreases as the temperature rises. The losses at 9,300 Mc/s for ferroxcube IVB and IVC are plotted in fig. 5 as a function of the temperature. It can be

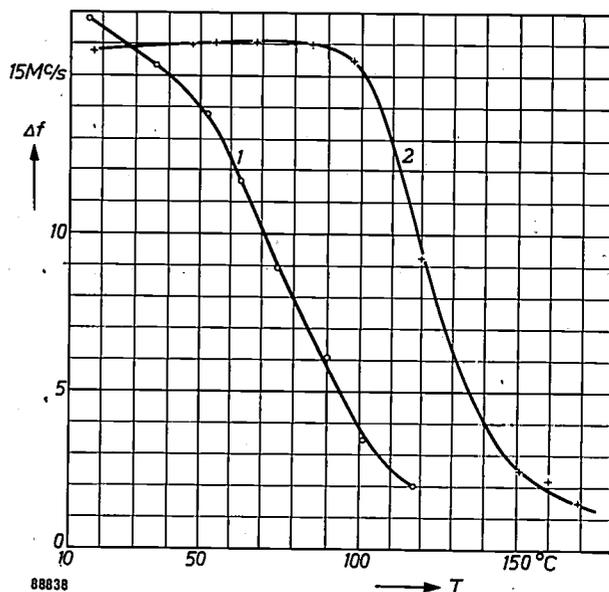


Fig. 5. The magnetic losses of ferroxcube IVB and IVC as functions of the temperature. The losses are expressed in terms of  $\Delta f$  (see fig. 3).

seen that the losses in ferroxcube IVB become small at 120 °C and that this occurs at 160 °C for ferroxcube IVC. We also know that  $M$  is approximately equal to  $2.6 \times 10^5$  A/m (3,300 oersted) in both these cases. Thus we find experimentally that the maximum resonant frequency  $\omega_{\max}$  is

$$\omega_{\max} \approx \gamma M,$$

in agreement with (5).

**Summary.** A centimetre wave transmitter can be modulated by controlled absorption of the power transmitted. An absorption of this type can be brought about by gyromagnetic losses in ferrites. An H.F. magnetic field exhibits losses in demagnetized ferrites over a large frequency range, e.g. from 10—10,000 Mc/s. If, however, the ferrite is magnetized by a weak external biasing field the spin directions are aligned and the losses are concentrated in a small frequency region in the neighbourhood of a certain high frequency; outside this region the losses are small. If the external biasing field at a given carrier frequency is allowed to vary in intensity, variations in losses arise. These variations may be used for absorption modulation, allowing modulation up to frequencies of 1.3 Mc/s. The carrier frequency for this modulation can in principle be chosen between 2,000—12,000 Mc/s provided a type of ferroxcube is used which has a suitable saturation magnetization value. The nature of the losses is discussed; they arise through resonance phenomena under the influence of crystal anisotropy fields and internal magnetic fields that can assume different values by demagnetization. The experimentally found maximum frequency at which losses occur can also be explained theoretically.

## THE MEASUREMENT OF RADIOACTIVE IODINE IN THE THYROID GLAND

621.387.4:546.15.02.131:616.44-07

In the diagnosis of disorders of the thyroid gland use is often made of the radioactive iodine isotope  $I^{131}$ . The assimilation of iodine from the food is a measure of the activity of the thyroid gland and, in principle, can be easily investigated by administering a dose containing the radioactive isotope and determining the radioactivity of the thyroid on a number of consecutive days with, say, a Geiger counter tube.

In practice, determination of the radioactivity of

The instrument shown<sup>1)</sup> was made feasible by the use of a Geiger counter tube (type no. 18503) whose active volume comprises practically the whole length of the tube (5 cm including the anode connection). Ten such tubes have been accommodated without difficulty in the collar<sup>2)</sup>. Moreover, the above-mentioned type of Geiger tube has the advantage of a low working voltage (less than 400 V) and a very flat plateau, so that it is not necessary to stabilize the voltage supply<sup>3)</sup>.



the thyroid gland is a matter of some difficulty, at least if the accuracy is to be within a few percent, since the position of the source of radiation with respect to the radiation detector, i.e. the position of the neck of the patient with respect to the counter, must be exactly the same for each measurement. The problem has been solved by using a ring of counter tubes connected in parallel, fitting like a collar round the patient's neck. Radiation, emitted in all directions by the thyroid, is then integrated (at least in one plane), and should the source be off-centre, a loss in one counter tube will be roughly compensated by the increased radiation received by the diametrically opposite tube.

The collar is calibrated by means of a phantom. If a bottle containing 5 microcuries of  $I^{131}$  — the order of magnitude of the  $I^{131}$  taken up by the thyroid during actual investigations — be placed in the centre of the collar, about 700 pulses/minute are recorded. The background caused by cosmic radia-

<sup>1)</sup> A. M. J. Jaspers and H. R. Marcuse, Local measurement of  $I^{131}$  uptake in the thyroid gland, *Rec. Trav. chim. Pays-Bas* **74**, 355-361, 1955 (No. 5). The apparatus has been in use for some time at the Municipal Hospital in The Hague.

<sup>2)</sup> Toroidal tubes, enclosing the whole neck, have been constructed for the same purpose; however, they are difficult to manufacture and for this reason have not been marketed; see W. Horst, *Fortschr. Röntgenstr.* **77**, 567, 1952.

<sup>3)</sup> See: N. Warmoltz, *Geiger-Müller counters*, *Philips techn. Rev.* **13**, 282-292, 1951/52.

tion is about 100 pulses/min. With a phantom of the patient's neck, constructed from meat, with a bone as the spinal column and a small rubber bag filled with a solution of radioactive iodine as the thyroid, the count obtained is about 10% higher than that from the iodine solution alone. One cause of this higher reading is the secondary radiation generated particularly in the spinal column, which

to about 2%. In order to reduce the statistical measuring error to roughly the same level, the pulses produced by 5  $\mu\text{C}$  must be counted over 15 minutes<sup>4</sup>). Usually the patient is given 20  $\mu\text{C}$  which results in a counting time of 3-5 minutes.

Further details as to the measurement and the diagnostic interpretation will not be gone into here, but it should be added that for accurate assessment



is better absorbed by the counter tubes. Such phenomena also result in a lower sensitivity to displacement of the source of radiation when the phantom is present; without a phantom 20% more pulses are obtained when the source of radiation is 4 cm off-centre, whereas with a phantom only 10% more pulses are recorded for a similar displacement.

The discrepancy of 10% is still inadmissible, but variations in the positioning of the collar round the neck can be kept very much less than 4 cm during the series of repeated measurements. This is done quite simply by means of fitting-rings which slot into the collar and fit round the neck of the patient; see photograph above. The error due to variation in the positioning of the collar is reduced in this way

of the amount of iodine in the thyroid, a correction must be applied to the results of the measurement for the amount of iodine in the blood flowing through the neck. This correction is found by recording the radioactivity with the collar around the thigh of the patient in conjunction with each measurement on the thyroid; from measurements on patients whose thyroids have been removed, the radioactivity measured at the thigh — where the count rate is 2-3 times the background — was found to be about  $\frac{2}{3}$  that of the bloodstream through the neck.

A. M. J. JASPERS.

<sup>4</sup>) See for example, Philips tech. Rev. 17, 206, 1955/1956 (No. 7-8).

## MECHANICAL PHENOMENA IN GRAMOPHONE PICK-UPS AT HIGH AUDIO FREQUENCIES

by J. B. S. M. KERSTENS.

534.851:681.842.081.3

---

*An article on a new gramophone pick-up will shortly appear in this review. As an introduction to this article we give here a study of the mechanics of the system pick-up - record, particularly at high audio frequencies. Extensive calculations on this question can be found in the literature; often, however, the physical aspect on the phenomena is liable to get lost and only after considerable simplification are useful results obtained. In this article the higher harmonics of the stylus motion are neglected at the outset, so that the mathematics does not cloud the general physical picture.*

---

Two phenomena are important in determining the frequency response of a pick-up in the higher audio range (about 5000 to 20 000 c/s). The first is the stylus-groove resonance, i.e. the resonance of the effective mass, thought of as concentrated in the stylus tip, and the stiffness of the walls of the groove in which the stylus rests. The second phenomenon is the loss of high notes owing to the elastic deformation brought about by the stylus tip in the walls of the groove, a loss that is greater the smaller are the radii of curvature of the groove walls, i.e. the higher the modulation frequency, the greater the amplitude and the nearer the pick-up is to the centre of the disc.

The theory of the elastic deformation of two curved bodies in mutual contact under the influence of a given force as given by Hertz<sup>1)</sup> can be applied to both phenomena. As early as 1941 Kornei<sup>2)</sup> developed a theory based on Hertz' work to explain the "playback loss" at higher frequencies.

On the basis outlined by Kornei, F. G. Miller<sup>3)</sup> developed a broadly conceived theory on the relations between groove and stylus. Apart from the deformation of the groove walls, he includes the distortion in the motion of the stylus when tracing a groove with perfectly rigid walls, in which he employs the calculations of Lewis and Hunt<sup>4)</sup>. This tracing distortion arises from the finite dimensions of the stylus tip, so that the curve traced by the stylus is not entirely identical with that cut into the record. In this further analysis Miller splits

up the complex motion of a stylus riding a sinusoidal groove into a fundamental component and higher harmonics; as regards the fundamental component he arrives at relatively simple formulae, from which are derived the phenomenon of stylus-groove resonance as well as the loss of high notes.

In order to gain a clearer picture of the physical phenomena, the higher harmonics will be disregarded here. We can thus confine ourselves to a greatly simplified derivation, keeping in mind, however, that this theory will not be entirely valid in extreme cases such as may occur sometimes in practice. This simplification also means that wherever we speak of deformations, these are considered to be of a purely elastic nature, according to Hooke's law. The highly complex phenomena occurring with plastic deformation do not lend themselves to a simple description<sup>5)</sup>.

In the following considerations of the pick-up as a mechanical system only the trajectory of the stylus tip will be studied; resonances of the mechanical transmission system and of the electrical circuit will be disregarded.

### Deformation of the groove walls

Before dealing with the deformation of the groove walls by the reproducing stylus, let us first consider the geometry of the groove and the forces exerted by the stylus on the groove walls. Let us assume that a sinusoidal signal

$$y = \hat{y} \sin \omega t$$

has been recorded, the  $y$ -axis being chosen in the plane of the record, normal to the unmodulated groove. During playback the angular frequency is determined by the "wavelength"  $\lambda$  registered in

1) Heinrich Hertz, *Gesammelte Werke*, III. Die Prinzipien der Mechanik. J. A. Barth, Leipzig 1894; S. Timoshenko, *Theory of elasticity*, McGraw-Hill, New York 1934.

2) O. Kornei, On the playback loss in the reproduction of phonograph records, *J. Soc. Mot. Pict. Engrs.* **37**, 569-590, 1941.

3) F. G. Miller, *Stylus-groove relations in phonograph records*, dissertation, Harvard 1950.

4) W. D. Lewis and F. V. Hunt, *J. Acoust. Soc. Amer.* **12**, 348-356, 1940/41. See also: J. A. Pierce and F. V. Hunt, *J. Acoust. Soc. Amer.* **10**, 14-28, 1938/39.

5) F. V. Hunt, On stylus wear and surface noise in phonograph playback systems, *J. Audio Eng. Soc.* **3**, 2-18, Jan. 1955.

the record, and by the linear velocity  $V_g$  of the groove, according to:

$$\omega = 2\pi \frac{V_g}{\lambda} \dots \dots \dots (1)$$

Fig. 1 shows part of such a sinusoidal groove; the stylus tip, represented by a sphere, is situated here just at the crest of the wave. At this spot both lateral amplitude and acceleration are greatest, so

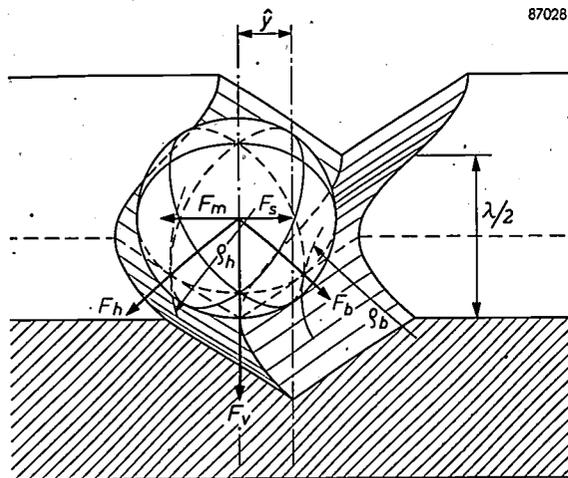


Fig. 1. The stylus tip, represented as a sphere, rests in the groove of a record, near the crest of the recorded sine wave. The walls of the groove are thought of as perfectly rigid.

that both stiffness force and inertia force will also be maximum. These forces can be represented respectively by

$$\hat{F}_s = s\hat{y}_a \text{ and } \hat{F}_m = m\omega^2\hat{y}_a, \dots \dots (2)$$

s being the stiffness and m the effective mass of the stylus system considered as acting through the centre of the stylus tip;  $\hat{y}_a$  is the amplitude of the actual stylus trajectory, which will generally differ from y, as will be shown later (suffix a for "actual", because the curve representing  $\hat{y}_a$  as a function of the frequency will be called the actual pick-up characteristic; see below).  $F_s - F_m$  is the nett lateral force  $F_l$  on the stylus tip.

Apart from this lateral force, a vertical force  $F_v$  acts on the stylus tip. If we disregard the "pinch" effect<sup>6)</sup> the latter force can be considered as constant.

In the foregoing no mention has been made of the damping of the pick-up. In general,  $F_l$  may be written as<sup>7)</sup>:

$$F_l = \left( \frac{s}{j\omega} + r + j\omega m \right) v \dots \dots \dots (3a)$$

<sup>6)</sup> See e.g. Philips tech. Rev. 13, 139, 1951/52.

<sup>7)</sup> If the method of complex numbers is used,  $F_l$  may also be expressed as the product of the mechanical impedance at the stylus tip and the velocity  $v = j\omega y_a$ :

$$F_l = \left( s + r \frac{d}{dt} + m \frac{d^2}{dt^2} \right) y_a, \dots \dots (3)$$

where r represents the damping of the pick-up. In fig. 2 the two forces  $F_l$  and  $F_v$  are represented again, with their components at right angles to the groove walls.

The groove angle  $2\beta$ , being very nearly  $90^\circ$  in practice, is assumed, for the sake of simplicity, to be exactly  $90^\circ$ . For the forces normal to the groove walls we then obtain<sup>8)</sup>

$$\begin{aligned} F_h &= \frac{1}{2}\sqrt{2} (F_v + F_l) \\ F_b &= \frac{1}{2}\sqrt{2} (F_v - F_l) \end{aligned} \dots \dots (4)$$

where the suffix h refers to the concave side of the groove and b to the convex side. These forces cause elastic deformations of the groove walls. The depth of the deformations is determined not only by these forces, but also by the nature of the materials and by the curvatures of the surfaces concerned. Evidently, a stylus tip with a small radius of curvature will penetrate deeper into the groove wall than one with a larger radius. It is likewise fairly obvious that a stylus tip of given radius will penetrate deeper into the convex than into the concave wall of the groove, since the contact area of the stylus with the latter is greater.

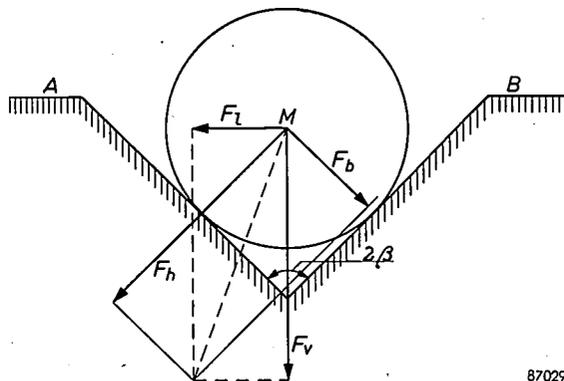


Fig. 2. Cross-section through the centre M of the stylus tip, for the situation shown in fig. 1. The groove-wall is concave at A, convex at B.

All this can be expressed in a formula given by Hertz for the elastic deformation of curved bodies in contact under a given force. For this special case the formula takes the form:

$$a = \left\{ \left( \frac{3}{2} \right)^2 F^2 \left( \frac{1 - \sigma^2}{E} \right)^2 \frac{1}{R} \left( 1 \pm \frac{R}{2q} \right) \right\}^{\frac{1}{3}} \dots \dots (5)$$

<sup>8)</sup> In fig. 2 the direction of  $F_l$  is to the left. This direction depends not only on the relative magnitude of  $F_s$  and  $F_m$  (the latter a function of frequency) but also on the damping, which introduces a phase angle between  $F_l$  and  $F_m$ . The effects of these quantities on the direction of  $F_l$  may be quantitatively examined by considering the mechanical impedance and the velocity as functions of the frequency; cf. eq. (3a) in footnote 7).

The symbol  $a$  represents the depth of penetration,  $F$  the force exerted by the stylus normal to the groove wall,  $E$  and  $\sigma$  respectively the Young's modulus and the Poisson's ratio of the record material (for both shellac and plastic records,  $\sigma \approx 0.35$ ),  $R$  the radius of the spherical tip of the stylus, and  $\rho$  the radius of curvature of the groove wall in a plane subtending an angle  $\beta$  with the horizontal plane. In (5) the plus-sign applies to a convex groove wall, the minus-sign to a concave one.

In Hertz's original formula a term appears with Young's modulus of the stylus material (at present generally sapphire or diamond) in the denominator. This modulus, however, is so much greater than that of the record material that this term can be neglected; in comparison with the record, the stylus can be regarded as perfectly rigid. The relevant values of  $E$  are roughly as follows<sup>9)</sup>:

Material	$E$ (newtons/m <sup>2</sup> )
Plastic . . . . .	$3 \times 10^9$
Shellac . . . . .	$6 \times 10^9$
Sapphire . . . . .	$500 \times 10^9$
Diamond . . . . .	$900 \times 10^9$

(The plastic referred to here is a copolymer of vinyl chloride and vinyl acetate<sup>10)</sup>, commonly used for long-playing records.)

Substituting the values of  $F$  given by (4) in (5) and grouping the constants together, we obtain for the penetration depths  $a_b$  and  $a_h$  in the convex and the concave side respectively:

$$a_{b,h} = \left\{ \left( \frac{3}{4} \right)^2 \left( \frac{F_v}{\sqrt{2}} \right)^2 \left( \frac{1-\sigma^2}{E} \right)^2 \frac{1}{R} \right\}^{\frac{1}{3}} \left\{ 1 \mp \frac{F_l}{F_v} \right\}^{\frac{2}{3}} \left\{ 1 \pm \frac{R}{2\rho} \right\} = a_0 \left\{ 1 \mp \frac{F_l}{F_v} \right\}^{\frac{2}{3}} \left\{ 1 \pm \frac{R}{2\rho} \right\} \quad (6)$$

where  $a_0$  represents the penetration depth in an unmodulated groove, since in this case  $F_l = 0$  and  $\rho = \infty$ .

It follows from (6) that the penetration depths can be reduced by increasing  $E$  (in this respect shellac is to be preferred to plastic), by increasing  $R$ , or by reducing  $F_v$ . Later in this article we shall consider the effect of  $F_l$  and  $\rho$ .

Fig. 3 shows a cross section of a stylus in an unmodulated groove indicating a typical static deformation. The value of  $a$  was derived from the values of  $F_v$ ,  $E$ ,  $\lambda$  and  $R$ . Fig. 4 is a perspective sketch of a stylus resting (stationary) in a modulated groove.

**Analysis of the stylus motion**

The trajectory of the stylus can be studied with the aid of the following imaginary experiment. The pick-up rests in a stationary groove, e.g. at the crest of a sine wave, as shown in fig. 1. Because the penetration in the convex groove wall is deeper than in the concave one, the tracing sphere rests such that the horizontal distance of its centre from the zero line (centre of unmodulated groove) is not  $\hat{y}$ , as would be the case with a perfectly rigid wall (fig. 5a) but smaller, say  $\hat{y}_s$  (s for the static situation considered) as in fig. 5b. If the stylus is exactly at the point where the groove passes through zero, the penetrations on either side will be equal, as the groove can be considered straight at this point. In this case, therefore, the horizontal projection of the centre coincides with the point of zero amplitude ( $N$  in fig. 5b).

Progressing accordingly for intermediate points, we arrive at a curve such as  $y_s$  in fig. 5b. A loss in

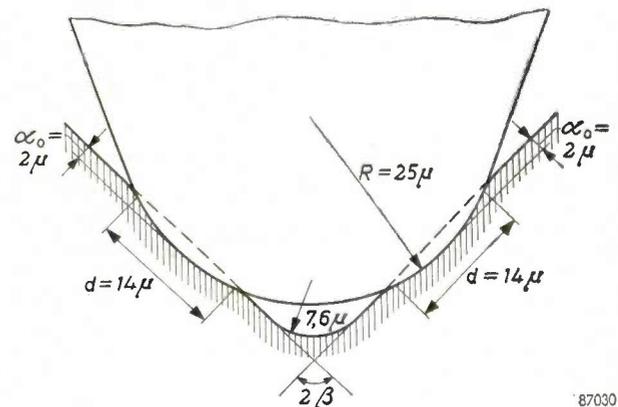


Fig. 3. Elastic penetration by the stylus in the walls of an unmodulated groove, drawn to scale for a co-polymer plastic record with  $E = 3.3 \times 10^9$  N/m<sup>2</sup> and  $\sigma = 0.35$ . Groove angle  $2\beta = 90^\circ$ . Apex angle of stylus  $45^\circ$ . Vertical force  $F_v = 0.1$  N ( $\approx 10$  g). Depth  $a_0$  of the penetration is  $2 \mu$ , diameter  $d$  of contact area is  $14 \mu$ .



Fig. 4. Perspective drawing, showing the static situation with the stylus resting in a modulated groove.

<sup>9)</sup> Values in M.K.S. units. The newton is the force necessary to produce an acceleration of 1m/sec<sup>2</sup> in a mass of 1 kg, i.e. the force exerted by the weight of 1/g kg, where  $g = 9.81$  m/sec<sup>2</sup>; this gives  $1 \text{ N} \approx 1/10$  kg weight. Hence to convert the values given to kg/mm<sup>2</sup> they should be multiplied by  $10^{-7}$ .

<sup>10)</sup> See Philips tech. Rev. 17, 108, 1955/56 (No. 4).

amplitude will be noted; this will be termed the *static tracing loss*. The amplitude  $\hat{\delta}_s$  of this loss is  $\hat{y} - \hat{y}_s$ .

As long as the record is stationary we are only concerned with a constant vertical force due to the stylus pressure, which can be resolved into two components normal to the groove walls. Although these components are equal in magnitude, it is owing to the opposite curvatures of the walls that they cause impressions of different shapes ( $O_1$  and  $O_2$  in fig. 5b). When the record is turning, additional lateral forces come into play. Let us assume that a sinusoidal signal of constant wavelength is recorded

force is then exerted on the stylus. Hence  $\delta_d$  becomes zero; the total loss is then equal to the static tracing loss. The frequency at which this happens is called the resonance frequency of the pick-up system. Since the external lateral force on the stylus is now zero, this resonance of the effective mass at the stylus tip and the stiffness of the pick-up system is the same as would occur with the stylus freely vibrating in air, outside a groove.

From  $F_m = F_s$ , the free resonant frequency is given (see eq. 2) by  $s\hat{y}_a = m\omega^2\hat{y}_a$ , or  $\omega = \sqrt{s/m}$ . This resonance does not appear in the actual

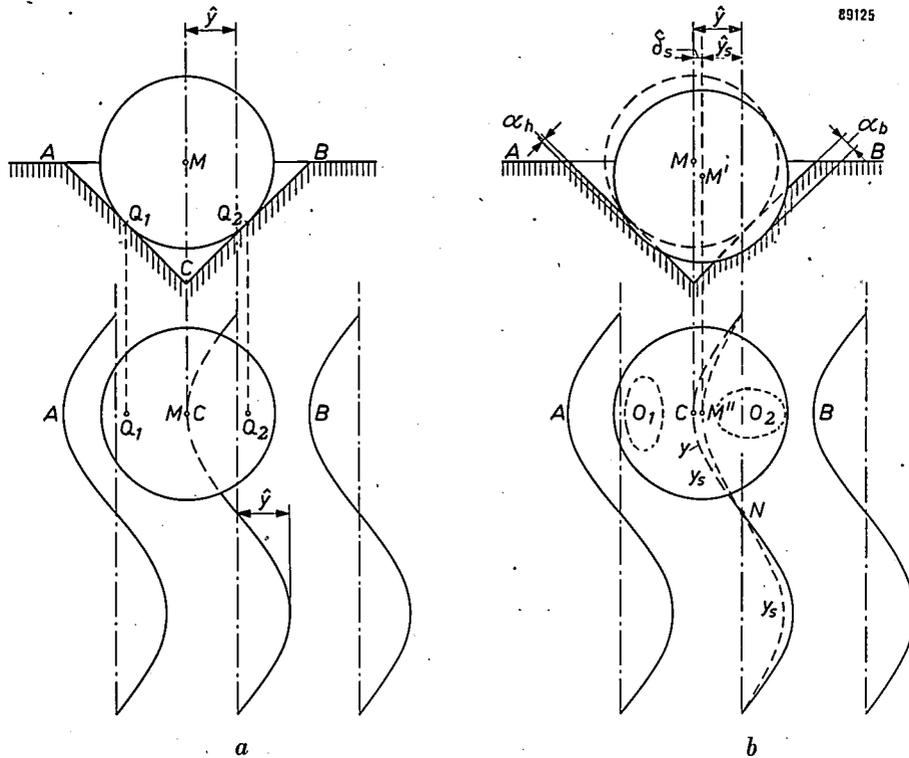


Fig. 5. a) Stylus in groove with perfectly rigid walls. Stylus and groove touch at only two points,  $Q_1$  and  $Q_2$ ; no horizontal displacement of the centre  $M$  of the stylus tip. b) Same for a groove with elastic walls. The points of contact of stylus and groove have now become areas ( $O_1$  and  $O_2$  shown exaggerated for clarity); the penetration  $\alpha_b$  at the convex side of the groove (B) is deeper than  $\alpha_h$ , and consequently  $M$  is shifted to  $M'$  over a distance whose horizontal projection  $CM''$  is  $\delta_s = \hat{y} - \hat{y}_s$  in the event that the stylus is situated at the crest of the sine. Instead of the recorded sine curve  $y$ , the stylus describes the curve  $y_s$  of smaller amplitude.

and that the record is now slowly started and gradually speeded up. According to equation (1) the frequency will then gradually increase. As long as the frequency is low, the stiffness force  $F_s$  will exceed the inertia force  $F_m$  and, apart from the earlier mentioned tracing loss, a *dynamic tracing loss*  $\delta_d$  will occur. The total tracing loss  $\delta_t$  is the sum of  $\delta_d$  and  $\delta_s$ .

As the frequency becomes higher,  $F_m$  increases until at a certain frequency  $F_m$  is just equal to the force  $F_s$  provided by the stiffness: no external lateral

pick-up characteristic; it occurs, for most pick-ups, at a frequency between about 1000 and 2000 c/s, which is considerably below the *stylus-groove* resonance (to be dealt with later) which has the major effect on the actual characteristic.

At frequencies above the resonance frequency of the pick-up system the force  $F_m$  will predominate and the dynamic tracing loss  $\delta_d$  will thus become negative, i.e. the total tracing loss will be less than the static loss. At a sufficiently high  $F_m$  (hence high frequency and/or large mass) the total loss

may even become negative, and there will therefore be some amplification of the signal. The amplification is greatest at the frequency at which the stylus-groove resonance (stylus-tip inertia and groove-wall stiffness) takes place. This resonance depends on the material of record and stylus, on the shape of groove and stylus and further on the vertical force.

Summarizing, we find that when an ordinary record (unlike our theoretical model modulated with varying wavelength) is played back two opposite effects occur: at higher frequencies the radii of curvature are smaller and therefore the static loss is large, while at the same time the inertia forces increases and thus the dynamic loss is reduced. (This is directly apparent from the signs in (6)). Which of the two effects is predominant depends upon several factors and will be discussed at the conclusion of this article, after dealing with each effect separately.

**Static tracing loss and cut-off frequency**

The static tracing loss was defined in the foregoing by  $\delta_s = y - y_s$ , where  $y$  represents the signal registered in the record and  $y_s$  the lateral deflection of the stylus tip. The static tracing loss increases according as the radius of curvature of the groove walls becomes smaller. At a given radius (and when the amplitude  $\ll$  radius of stylus point, as it always is for the frequencies under consideration), the loss can even equal the original signal:  $y = \delta_s$ . Then  $y_s = 0$ ; in other words, the stylus describes an unmodulated trajectory and the pick-up does not produce any voltage. We shall evaluate this case for the peak of a sine wave; at this point the deformation is largest and the calculation is simplest.

The following relation exists between the absolute value  $|e|$  of the radius of curvature at the peaks and the wavelength or frequency:

$$|e| = \frac{V_g^2 \sqrt{2}}{\hat{y} \omega^2} = \frac{\lambda^2 \sqrt{2}}{4\pi^2 \hat{y}} \dots \dots (7)$$

where  $V_g$  is the velocity of the stylus relative to the unmodulated groove.

The general expression for the radius of curvature of a curve, in Cartesian coordinates, is

$$e = \frac{\left\{ 1 + \left( \frac{dy}{dx} \right)^2 \right\}^{\frac{3}{2}}}{\frac{d^2y}{dx^2}}$$

In the present problem, the  $x$  coordinate is the direction of the unmodulated groove and the  $y$  coordinate is the displacement of the stylus from this axis. At a wave peak

$dy/dx = 0$ , hence

$$e = \frac{1}{d^2y/dx^2}$$

Since  $y = \hat{y} \sin \omega t = \hat{y} \sin (\omega x/V_g)$ , we have, at the peaks,  $d^2y/dx^2 = -\hat{y} \omega^2/V_g^2 = -4\pi^2 \hat{y}/\lambda^2$ . The factor  $\sqrt{2}$  is introduced in (7) because the radius of curvature considered is in a plane at an angle of  $45^\circ$  with the horizontal plane.

The frequency and wavelength corresponding to the radius of curvature at which the stylus no longer moves laterally, and at which the loss is therefore just as large as the original signal recorded, are called the *cut-off frequency* and the *cut-off wavelength*. At the cut-off frequency,  $F_1$  is zero, since the lateral deflection is zero (cf. eq. (3)). However, we shall assume here that  $F_1$  is zero throughout the whole frequency range in order to distinguish the static tracing loss from phenomena due to the action of an additional lateral force associated with the dynamic tracing loss. The formulae (6) for the penetration can then be expressed in the following simplified form:

Convex side:  $a_b = a_0 \left( 1 + \frac{R}{2\rho} \right)^{\frac{1}{2}}$ ,

Concave side:  $a_h = a_0 \left( 1 - \frac{R}{2\rho} \right)^{\frac{1}{2}}$ .

Now, by geometry,

$$\hat{\delta}_s = \frac{a_b - a_h}{2}$$

Expanding the above expressions for  $a_b$  and  $a_h$  (neglecting higher powers of  $R/2\rho$ ) and substituting we get

$$\hat{\delta}_s \approx \frac{a_0}{\sqrt{2}} \times \frac{2}{3} \frac{R}{2\rho}$$

Using (7) this becomes

$$\hat{\delta}_s = \frac{a_0 R}{6 V_g^2} \omega^2 \hat{y}$$

Hence

$$\hat{y}_s = \hat{y} - \hat{\delta}_s = \hat{y} \left( 1 - \frac{a_0 R}{6 V_g^2} \omega^2 \right) \dots (8)$$

The cut-off frequency  $f_{co} = \omega_{co}/2\pi$  and the cut-off wavelength  $\lambda_{co}$  are the values at which  $\hat{y}_s$  is zero, and hence

$$\omega_{co} = V_g \sqrt{\frac{6}{a_0 R}} \text{ and } \lambda_{co} = \frac{2\pi V_g}{\omega_{co}} = 2\pi \sqrt{\frac{a_0 R}{6}} \dots (9)$$

We now obtain for  $\hat{y}_s$  the following quadratic expressions in terms of  $\omega/\omega_{co}$  and of  $\lambda_{co}/\lambda$ :

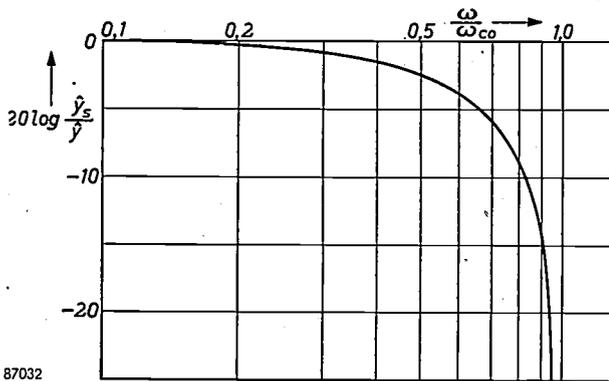
$$\hat{y}_s = \hat{y} \left\{ 1 - \left( \frac{\omega}{\omega_{co}} \right)^2 \right\} = \hat{y} \left\{ 1 - \left( \frac{\lambda_{co}}{\lambda} \right)^2 \right\} \dots (10)$$

The former expression is plotted in *fig. 6* (on logarithmic scale).

Since  $\omega$  and  $\omega_{co}$  are equally dependent upon the groove velocity  $V_g$  (equ. (1) and (9)), the effect of this velocity is not expressed in *fig. 6*. The fact that this effect is considerable can be readily demonstrated with the aid of a numerical example. Let us assume that  $F_v = 0.1$  N ( $\approx 10$  grams),  $E = 3.3 \times 10^9$  N/m<sup>2</sup>,  $\sigma = 0.35$  and  $R = 25 \times 10^{-6}$  m; we then find from (6):  $a_0 = 2 \times 10^{-6}$  m. The cut-off wavelength is then  $\lambda_{co} = 2\pi\sqrt{a_0 R/6} = 18\mu$ , and hence the cut-off frequency is

$$f_{co} = V_g/\lambda_{co} = 55\,000 V_g \text{ c/s, . . . (11)}$$

$V_g$  being expressed in metres per second. On the outside of a  $33\frac{1}{3}$  r.p.m. 12'' record,  $V_g = 0.50$  m/s and  $f_{co}$  is consequently 27 500 c/s; on the inside, however,  $V_g = 0.20$  m/s and  $f_{co}$  therefore only 11 000 c/s. Under such conditions, therefore, reproduction above 11 000 c/s is impossible near the



*Fig. 6.* The ratio  $\hat{y}_s/\hat{y} = 1 - (\omega/\omega_{co})^2$  as a function of  $\omega/\omega_{co}$  (static characteristic, cf. *fig. 10*); both are plotted on logarithmic scales.

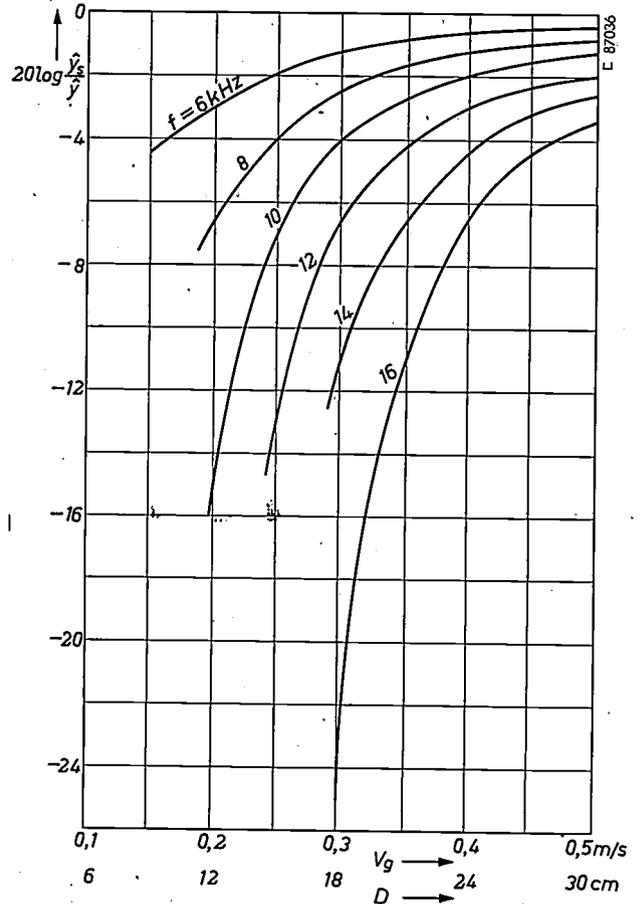
end of the record, and considerable attenuation occurs even below 11 000 c/s. (This attenuation is more or less compensated when the dynamic tracing loss becomes negative, but it is impossible to profit by this effect above the cut-off frequency).

In *fig. 7*,  $20 \log \hat{y}_s/\hat{y}$  is plotted as a function of the groove velocity of a  $33\frac{1}{3}$  r.p.m. record under the above-mentioned conditions, for various frequencies (eq. 8). It can be seen that the reproduction of the high notes is decidedly poorer on the inside than on the outside of the record.

**Dynamic pick-up characteristic, dynamic tracing loss and resonance frequency**

The dynamic pick-up characteristic is the characteristic after elimination of the static tracing loss dealt with in the foregoing (thus including the dynamic tracing loss).

The dynamic pick-up characteristic can be derived in a simple way by means of an electromechanical analogy<sup>11</sup>). *Fig. 8a* is the equivalent diagram of the mechanical system. The signal  $\hat{y} \sin \omega t$  is transferred via the elastic groove wall of stiffness  $S$  to the effective mass  $m$  concentrated in the stylus tip, the deflection  $y_d$  (suffix  $d$  for dynamic) of which we wish to find. The stiffness and the damping of the pick-up are denoted by  $s$  and  $r$  respectively.



*Fig. 7.*  $\hat{y}_s/\hat{y}$  plotted logarithmically as a function of the groove velocity  $V_g$ , for a  $33\frac{1}{3}$  r.p.m. record, for various frequencies  $f$ . For a given r.p.m.,  $V_g$  is proportional to the distance of the groove from the axis of rotation, i.e. to the groove diameter  $D$ . The ordinate is therefore also expressed in terms of  $D$ .

The electrical analogue of this system is shown in *fig. 8b*. The motion of the stylus tip can easily be evaluated from this diagram. We find:

$$\hat{y}_d = \frac{\hat{y}}{\sqrt{\left(1 + \frac{s - m\omega^2}{S}\right)^2 + \left(\frac{r\omega}{S}\right)^2}} \quad (12)$$

<sup>11</sup>) For the explanation and the application of electromechanical analogies, reference is made to the exhaustive literature on this subject, of which we particularly recommend: B. Gehlshøj, *Electro-mechanical and electro-acoustical analogies*, Ingenørvidenskabelige Skrifter (no. 1), Akad. tekn. Videnskaber, Copenhagen 1947.

By introducing the natural angular frequency of the stylus-groove system we obtain:

$$\omega_0^2 = \frac{S + s}{m} \quad (13)$$

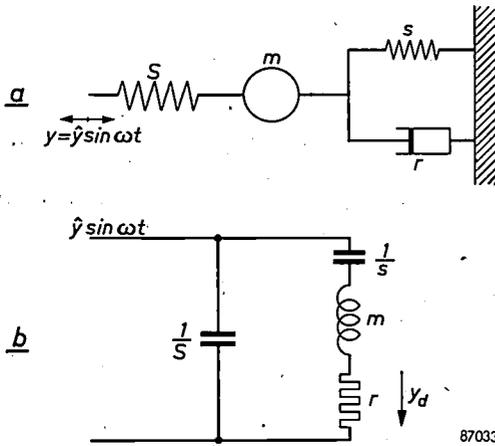


Fig. 8. a) Model of the pick-up mechanical system.  $m$  effective mass and  $s$  stiffness at the stylus tip,  $r$  damping,  $S$  stiffness of the groove wall.  $y_d$  is the displacement of the stylus tip resulting from an input signal  $y$ . b) Electrical analogue of (a).

Remembering that  $s$  is many times smaller than  $S$ <sup>12</sup>), so that  $\omega_0^2 \approx S/m$ , and defining the relative damping coefficient as

$$\varepsilon = \frac{r}{m\omega_0}$$

we can finally put:

$$\hat{y}_d = \frac{\hat{y}}{\sqrt{\left\{1 - \left(\frac{\omega}{\omega_0}\right)^2\right\}^2 + \varepsilon^2 \left(\frac{\omega}{\omega_0}\right)^2}} \quad (14)$$

This formula represents the dynamic pick-up characteristic. The variation of  $\hat{y}_d$  as a function of the frequency for various values of the damping is shown in fig. 9.

The frequency at which  $\hat{y}_d$  assumes the maximum value is the stylus-groove resonance frequency  $f_{res} = \omega_{res}/2\pi$ . By putting  $d\hat{y}_d/d\omega = 0$ , we find

$$\omega_{res}^2 = \omega_0^2 \left(1 - \frac{1}{2}\varepsilon^2\right) = \omega_0^2 - \frac{r^2}{2m^2}$$

The fact that the resonance frequency is lower than the natural frequency is due to the presence of damping; if the damping is small, however, the difference is negligible.

In the foregoing we have introduced the stiffness  $S$  of the groove walls. The value of  $S$  can be easily

related to  $F_v$  and  $\alpha_0$  if we take a simplified case<sup>13</sup>). If the stylus rode in an unmodulated groove and was driven by an external lateral force  $F_l$ , the depths of penetration would be given by a simplified form of (6) since  $\rho = \infty$ , i.e:

$$\alpha_{b,h} = \alpha_0 \left(1 \mp \frac{F_l}{F_v}\right)^{\frac{2}{3}}$$

The difference between the recorded and the traced signal,  $y - y_d$  is what we have called in the foregoing the dynamic tracing loss  $\delta_d$ ;  $y_d = y - \delta_d$  is just a simplified expression of (12). The magnitude of the loss  $\delta_d$  is given by expanding the above expressions for  $\alpha_b$  and  $\alpha_h$  (neglecting higher powers of  $F_l/F_v$ ), subtracting and dividing by  $\sqrt{2}$ :

$$\delta_d = \frac{\alpha_b - \alpha_h}{\sqrt{2}} \approx \frac{\alpha_0}{\sqrt{2}} \times \frac{4}{3} \frac{F_l}{F_v}$$

The groove wall stiffness  $S$  is defined as the ratio of the force  $F_l$  to the deflection  $\delta_d$  due to that force in the presence of the vertical force  $F_v$ . Hence

$$S = \frac{F_l}{\delta_d} = \frac{F_l}{\frac{\alpha_0}{\sqrt{2}} \frac{4}{3} \frac{F_l}{F_v}} = \frac{3}{2} \frac{F_v}{\alpha_0 \sqrt{2}} \quad (15)$$

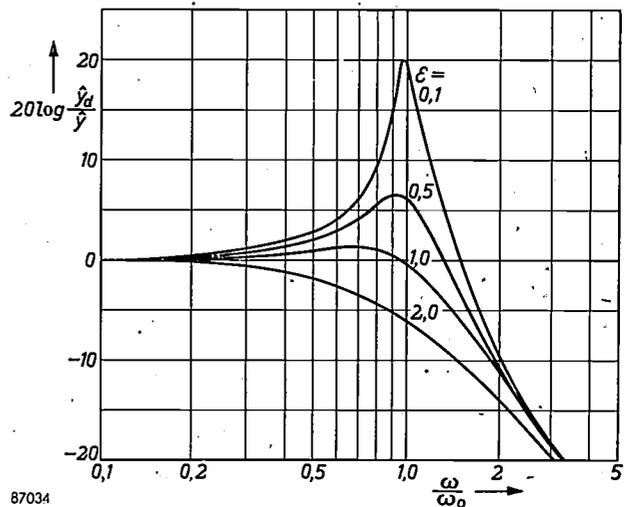


Fig. 9. Dynamic pick-up characteristic  $\hat{y}_d = f(\omega/\omega_0)$ , for various values of the relative damping coefficient  $\varepsilon$  (see eq. (14)). Provided that  $\varepsilon$  is not too large a peak occurs at a frequency (the stylus-groove resonance frequency) which lies further below the stylus-groove natural frequency  $\omega_0/2\pi$  the stronger the damping. Both scales are logarithmic.

**The actual pick-up characteristic**

The actual pick-up characteristic gives the actual stylus trajectory  $\hat{y}_a$  as a function of the frequency. It consists of a combination of the dynamic pick-up characteristic and the static tracing loss.

<sup>12</sup>) For ordinary pick-ups,  $s = 0.02$  to  $1$  kN/m, and for the groove wall  $S = 40$  to  $80$  kN/m (depending on the material used, on the vertical force  $F_v$  and on the radius  $R$  of the stylus tip).

<sup>13</sup>) It can be shown mathematically that the result obtained here is equivalent to that obtained when the actual conditions are taken into account.

The equation of the actual characteristic can be derived as follows. Consider once again the expression (14) for the dynamic characteristic, keeping in mind that the static loss should, in fact, also be taken into account, so that in (14)  $\hat{y}$  should be replaced by  $\hat{y}_s$ , i.e.

$$\hat{y}_a = \frac{\hat{y}_s}{\hat{y}} \cdot \hat{y}_d.$$

Dividing this by  $\hat{y}$  and taking logarithms, we obtain:

$$\log \frac{\hat{y}_a}{\hat{y}} = \log \frac{\hat{y}_s}{\hat{y}} + \log \frac{\hat{y}_d}{\hat{y}} \dots (16)$$

Hence if the static and the dynamic characteristics are plotted logarithmically on the same frequency scale, we obtain, as is evident from (16), the actual characteristic by adding the two functions. Alternatively, using the expression  $y_d = y - \delta_d$ , we can obtain the actual characteristic in an analogous manner by substituting in it  $y - \delta_s$  for  $y$ . The result is then:

$$y_a = y - \delta_s - \delta_d = y - (\delta_s + \delta_d) = y - \delta_t,$$

where  $\delta_t$  is the total loss.

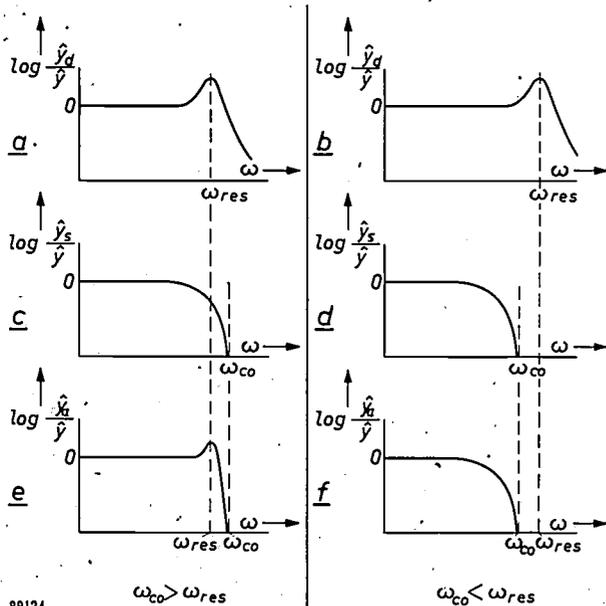


Fig. 10. (a) and (b) dynamic characteristic  $\log \hat{y}_d/\hat{y}$ , (c) and (d) static characteristic  $\log \hat{y}_s/\hat{y}$ , and (e) and (f) actual characteristics  $\log \hat{y}_a/\hat{y} = \log \hat{y}_d/\hat{y} + \log \hat{y}_s/\hat{y}$ ; (a), (c) and (e) refer to  $\omega_{co} > \omega_{res}$ ; (b), (d) and (f) refer to  $\omega_{co} < \omega_{res}$ .

$\log \hat{y}_d/\hat{y}$ ,  $\log \hat{y}_s/\hat{y}$  and their sum  $\log \hat{y}_a/\hat{y}$ , are plotted in fig. 10, on the left for the case of  $\omega_{co} > \omega_{res}$ , and on the right for  $\omega_{co} < \omega_{res}$ .

In the former case the drop due to the static loss below the resonance frequency is reduced by the resonance peak in  $\hat{y}_d$ ; there remains a resonance

peak in  $\hat{y}_a$ , although less pronounced than in  $\hat{y}_d$ , whilst the in region beyond the resonance frequency the cut-off is now sharper. In the second case ( $\omega_{co} < \omega_{res}$ ) the actual characteristic is virtually identical with the static one.

**The practical importance of the various properties of pick-up and record**

Consider the resonant angular frequency  $\omega_0$  of the stylus-groove system and the cut-off wavelength  $\lambda_0$ . For the former quantity we take (13) in the approximate form  $\omega_0^2 \approx S/m$ , substitute for  $S$  from (15) and then insert the value of  $a_0$  defined by equation (6). For  $\lambda_{co}$ , we simply insert the value of  $\alpha_0$  from (6) in equation (9). We then obtain:

$$\omega_0 = \left\{ \frac{6 F_v R}{\sqrt{2}} \frac{E^2}{m^3 (1 - \sigma^2)^2} \right\}^{\frac{1}{6}} \dots (17a)$$

$$\lambda_{co} = \left\{ \frac{\pi^3}{\sqrt{12}} F_v R \frac{1 - \sigma^2}{E} \right\}^{\frac{1}{3}} \dots (17b)$$

$\omega_0$  and  $\lambda_{co}$  are quantities dependent only on the physical properties of the pick-up, the properties of the record material, and the vertical force  $F_v$ . They are therefore independent of the groove velocity and of the amplitude of the signal. For a given pick-up and a given record,  $\omega_0$  and  $\lambda_{co}$  are to be regarded as system constants, which can be varied only by changing  $F_v$ . The cut-off frequency, however, as we have shown earlier, is determined by the relation  $\omega_{co} = 2\pi V_g/\lambda_{co}$  and is consequently directly proportional to the groove velocity:  $\omega_{co}$  will thus be higher on the outside than on the inside of the record. The position of the stylus on the record will therefore be one of the factors deciding whether  $\omega_{co}$  is greater or smaller than  $\omega_{res}$  (in most cases  $\omega_{res}$  is virtually equal to  $\omega_0$ ). From the above the following conclusions may be drawn:

- 1) A reduction of the effective mass  $m$  at the stylus tip has no effect on  $\lambda_{co}$  but it does raise the natural frequency  $\omega_0$  and consequently improves the reproduction of the high notes, provided that  $\omega_0 < \omega_{co}$ .
- 2) The stiffness of the pick-up does not appear in (17a or b) and therefore affects neither  $\lambda_{co}$  nor  $\omega_0$ . This is valid only for the high frequencies discussed here;  $s$  is obviously of importance at the lower frequencies.
- 3) The damping  $r$  of the pick-up determines the height of the resonance peak; it does so, however, in conjunction with the damping of the record material. The latter factor does not appear in the above considerations owing to the simplification introduced.

- 4) Increasing the Young's modulus  $E$  of the record material produces an increase in  $\omega_0$  and a reduction of  $\lambda_{co}$ . As mentioned in connection with the groove deformation, a shellac record, whose Young's modulus is about twice that of a plastic record, will be more favourable in this respect. Owing to other factors, however, such as the reduced noise and the unbreakability, the plastic material is nevertheless to be preferred (cf. the articles quoted in <sup>9)</sup> and <sup>10)</sup>).
- 5) The radius of curvature  $R$  of the stylus tip is  $25 \mu$  for long-playing records and  $75 \mu$  for normal (78 r.p.m.) records. This greater radius causes a higher  $\omega_0$  but also a greater  $\lambda_{co}$  and, therefore, a greater static tracing loss if the same record material is used. Normal records, however, are usually made of shellac, which, as mentioned in 4) above, has a greater Young's modulus than plastic, and this partly compensates for the greater loss due to the greater radius of curvature of the stylus tip. According to (17b)  $\lambda_{co}$  is proportional to  $(R/E)^{\frac{1}{3}}$ , so that the cut-off wavelength  $\lambda_{co}$  for shellac ( $R = 75 \mu$ ,  $E = 6 \times 10^9 \text{ N/m}^2$ ) will be a factor of

$$\left(\frac{75}{25} \times \frac{3.3}{6}\right)^{\frac{1}{3}} = 1.18$$

greater than for co-polymer plastic ( $R = 25 \mu$ ,  $E = 3.3 \times 10^9 \text{ N/m}^2$ ). On the smallest diameter (9.5 cm) of a 78 r.p.m. record the groove velocity  $V_g$  is approx 1.85 times greater than on the smallest diameter (12.5 cm) of a  $33\frac{1}{3}$  r.p.m. record, so that according to equation (1) the

cut-off frequency in the former case will be higher by a factor of  $1.85/1.18 \approx 1.6$  than in the latter case.

- 6) The vertical force  $F_v$ , to which the stylus pressure is proportional, also affects  $\omega_0$  and  $\lambda_{co}$  as shown by (17): a reduction of the stylus pressure reduces  $\lambda_{co}$  and together with it the tracing loss, but  $\omega_0$  is then also reduced.

In a later article it will be shown how the above considerations have been applied in the design of a new pick-up.

**Summary.** As an introduction to an article on a new gramophone pick-up the mechanics of the system pick-up — gramophone record are considered. The stylus (sapphire, diamond) whose spherical tip rides in the groove of the record, causes elastic deformation of the groove walls. The convex wall is deformed to a greater depth than the concave wall. This causes a loss in the stylus deflection, the static tracing loss. This loss is greater the smaller the radius of curvature of the groove, i.e. the higher the frequency, the larger the amplitude and the nearer the stylus is to the centre of the disc. At a given frequency — the cut-off frequency — this loss is equal to the recorded signal, so that the stylus tip does not vibrate at all; this loss becomes apparent even considerably below the cut-off frequency.

A second effect — the dynamic tracing loss — has likewise the nature of a loss as long as the stiffness force of the pick-up exceeds the inertia force on the stylus tip; this is the case at frequencies below the free resonance frequency of the pick-up system. At higher frequencies this loss becomes negative and thus acts in opposition to the static loss. The total loss (the sum of the static and the dynamic losses) may even become negative. This is liable to happen particularly in the neighbourhood of the frequency at which resonance of the groove-wall stiffness and the effective mass at the stylus tip occurs; the resonance curve of this system is called the dynamic pick-up characteristic. A combination of the latter with the static characteristic produces the actual pick-up characteristic. According to whether the cut-off frequency lies higher or lower than the stylus-groove resonance frequency, the actual characteristic will either show a slight peak or no peak at all. The article is concluded with a discussion of the practical significance of the various properties of pick-up and record.

## ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk \* can be obtained free of charge upon application to be Philips Research laboratory, Eindhoven, Netherlands.

- 2296:** P. W. L. Iterson and H. A. Teunissen: General outline of the development of transmitters for television and frequency modulation (T. Ned. Radiogenootschap 20, 207-222, 1955, No. 3).

Description of some TV and FM transmitters together with their auxiliary equipment developed in recent years by Philips Telecommunications Industries. Some details are given of design and properties. Special features are the use of high power tetrodes, a special FM modulation circuit and unit

construction which lead to simple installations of high performance. An FM transmitter is described which works without supervision and switches over automatically to reserve equipment in the event of a failure.

- 2297:** H. O. Huisman, R. van der Veen and J. Meltzer: A new acaricide, 2,4,5,4'-tetrachlorodiphenylsulphone (Nature 176, 515-516, 10 September 1955).

Note on the action of the above-mentioned acaricide. It acts on the eggs and larvae of a number

of mites but is ineffective against the fully-grown animals. The action is very selective: only mites are sensitive. Insects show no reaction and the acaricide is not poisonous to mammals.

**2298:** H. Groendijk and K. S. Knol: Characterization of the noise of tubes and transistors by four measurable quantities (T. Ned. Radio-genootschap **20**, 243-256, 1955, No. 4).

The noise of a neutralized triode is first calculated by investigating the mechanism of noise production and then by regarding the triode as a linear four-terminal network. It appears, from the first method, that the physical quantities connected with the generation of noise may not readily be determined. The second method shows, however, that it is still possible to characterize the noise by four measurable quantities and, once these are known, to calculate the noise factor. This holds for any linear four-terminal network, triodes, pentodes, transistors, etc. In general, no simple relationship exists between these quantities and the physical properties giving rise to noise. In the case of a triode the four quantities depend on frequency in a simple way. See also these abstracts No. R 277.

**2299:** J. A. Lely: Darstellung von Einkristallen von Siliciumcarbid und Beherrschung von Art und Menge der eingebauten Verunreinigungen (Ber. Deutsch. Keram. Gesellsch. **32**, 229-231, 1955, No. 8). (Preparation of single crystals of silicon carbide and control of type and quantity of lattice impurities; in German).

See these abstracts No. 2267.

**2300:** A. C. van Dorsten and J. H. Spaa: Een hoogspanningsgenerator in druktank voor neutronenproductie (Ned. T. Natuurk. **21**, 213-218, 1955, No. 9). (A pressurized high-voltage generator for neutron production; in Dutch).

Short description of equipment for producing a neutron flux of  $10^{10}$  fast neutrons per second by the bombardment of a heavy ice target with deuterons. See Philips tech. Rev. **17**, 109-111, 1955/56.

**2301:** B. Combée, J. Houtman and A. Recourt: A sealed-off X-ray tube for contact-micro-radiography (Brit. J. Radiol. **28**, 537-542, October 1955).

Short description of a new sealed-off X-ray tube for contact microradiography using a  $50 \mu$  Be window (see also Philips tech. Rev. **17**, 45-46, 1955/56). Calculations and measurements are given concerning the properties of photographic emulsions

in the interesting range of wavelengths between 2-12 Å.

**2302:** M. Gevers: Measurement of dielectric and magnetic properties of solids at microwave frequencies (Proc. Symposium Precision Electrical Measurements, National Physical Laboratory, Teddington, Nov. 1954).

Description of two of the methods used for measuring dielectric and magnetic properties of solids at cm wavelengths in the Philips Research Laboratories, Eindhoven. At 10 cm a resonance method is used and at 3 cm a method based on standing waves. Special attention is paid to the sources of errors and the accuracy of the measurements on low loss materials.

**2303:** J. W. L. Köhler: Machines frigorifiques à gaz froid (Annexe 1955-2, Suppl. Bull. Inst. Int. Froid, pp. 13-23).

See Philips tech. Rev. **16**, 69-78 and 105-115, 1954/55. Special attention is paid in this article to the properties of the regenerator which include one condition not hitherto recognized.

**2304:** J. Volger: Pertes diélectriques de quelques verres et de substances similaires, mesurée aux basses températures (Annexe 1955-2, Suppl. Bull. Inst. Int. Froid, pp. 89-98).

The dielectric losses of quartz and various silicate glasses have been studied at low frequencies. Round about 50 °K a new type of loss — so-called deformation loss — appears. The relation between these phenomena and the glass structure is discussed.

**2305:** G. Diemer: Transit-time functions of a dynatron oscillator. (Appl. sci. Res. B 4, 457-463, 1955).

The transit-time functions of the dynode admittance are discussed for a new type of dynatron-oscillator, in which the primary electrons are directed obliquely towards the dynode.

**2306:** J. A. Kok and M. M. G. Corbey: Breakdown of liquid insulating and dielectric material (Appl. sci. Res. B 4, 474-475, No. 6).

Preliminary note on experiments and theory of long-term breakdown of liquid dielectrics at low frequencies.

**2307:** J. Meltzer: Het onderzoek van acariciden en spintoviciden in het laboratorium (Tijdschrift Plantenziekten **61**, 130-142, 1955). (The investigation of acaricides and red spider ovicides in the laboratory; in Dutch, summary in English).

Description of laboratory methods for the evaluation of acaricidal and ovicidal action of a large number of substances. The substances concerned fall into three groups: those active on adult mites, those active on eggs and immature mobile stages and those active on all three stages. The results are summarized in 13 tables.

**2308:** L. J. Tummers: Der Einfluss von Minoritätsträger-Injektionen auf das Verhalten von Leistungstransistoren (Nachr.-techn. Fachber., 1955, No. 1, pp. 31-32). (The effect of minority carrier injection on the behaviour of power transistors; in German).

The emitter efficiency of a p-n-p alloy transistor with indium as the alloying material can be increased by adding a small amount of gallium or aluminium to the emitter indium. This improved emitter efficiency causes a more gradual fall off of the current amplification at increasing emitter currents and furthermore makes it possible to produce transistors with lower base resistance.

**2309:** J. C. van Vessem and Th. W. Willemse: Konstruktive Überlegungen beim Entwurf eines Leistungstransistors (Nachr. techn. Fachber. 33-36, 1955, No. 1). (Design considerations concerning power transistors; in German).

Design of a power transistor for a dissipation of 3 watts and description of the considerations which lead to this design.

**2310:** P. F. Moleman: Thermische Probleme bei der Verwendung von Schichttransistoren (Nachr. Techn. Fachber. 37-39, 1955, No. 1). (Thermal problems in the application of junction transistors; in German).

Study of the properties of a well-known circuit for the temperature stabilization of a transistor. When this circuit is used in a push-pull arrangement it is necessary to take account of the effect of temperature not only on the collector leakage current but also on the emitter-base potential difference.

**2311:** G. H. Jonker and H. R. Kruyt: Flocculation and recrystallization in freshly prepared silver bromide sols (Discussions Faraday Soc. No. 18, 170-190, 1954).

The ageing of diluted silver bromide sols has been studied by extinction measurements. The sols initially contain very small particles which flocculate to a certain degree, depending on the excess concentration of  $\text{Ag}^+$  or  $\text{Br}^-$  ions. In very diluted sols

a new type of coarsening is found in the region of normally stable negative sols. By studying the influence of electrolytes, protecting colloids and temperature it is found that this new type of coarsening consists of two stages: first a recrystallization takes place to rather large crystals and then still larger particles are formed by oriented flocculation of the ideally formed crystals. The same effect has been found in sols of silver chloride at higher concentrations and in silver iodide at lower concentrations.

**2312:** F. A. Kröger and D. de Nobel: Preparation and properties of CdTe single crystals (J. Electronics 1, 190-202, 1955).

Single crystals of CdTe were prepared and purified using the method described in abstract 2313 (see below). By re-heating below the melting point under suitable atmospheres the stoichiometry and hence the electrical properties were modified. Both *p* and *n* type conductivity have been obtained. Measurements of Hall effect, the thermoelectric power and the resistance give values of the mobility and the effective mass of electrons and holes and an idea of the energies of donor and acceptor levels.

**2313:** J. van der Boomgaard, F. A. Kröger and H. J. Vink: Zone melting of decomposing solids (J. Electronics 1, 212-217, 1955).

Zone melting of substances which tend to decompose on heating can be accomplished by heating in an atmosphere of the component formed on decomposition. By means of the pressure of this component it is possible to control the composition of the crystals obtained from the melt. As an example, the growth of a PbS crystal with a *p-n* junction is described.

**2314:** R. Vermeulen: Stereophonie und Stereomachhall (Musik-Raumgestaltung-Elektroakustik, 1955, Ars vira Verlag, Mainz). (Stereophony and artificial reverberation; in German).

See Philips techn. Rev. 17, 258-266, 1955/1956.

**2315:** H. Koelmans and J. Th. G. Overbeek: Stability and electrophoretic deposition of suspensions in non-aqueous media (Disc. Faraday Soc. 1-12, 1954).

The stability of suspensions in solvents of very low polarity is treated in part 1 of this paper. Theoretical considerations lead to the conclusion, that quite modest electric charges and  $\zeta$ -potentials are sufficient to stabilize suspensions of coarse particles ( $> 1 \mu$ ) whereas hardly any stabilization

can be expected from adsorbed layers of non-ionized long-chain molecules. Experiments on the settling times of suspensions of a number of solids in xylene confirm that only ionized surface-active substances give rise to stability. Long-chain compounds that do not increase the conductance of the xylene, do not give rise to a sufficient  $\zeta$ -potential of the particles and do not improve the stability very much. In part 2 electro-deposition from suspensions in polar organic media is investigated. It is shown that the particles are accumulated near the electrode by the applied field, but that the formation of an adhering deposit is caused by flocculation introduced by the electrolyte formed by the electrode reaction.

**2316:** A. G. Th. Becking, H. Groendijk and K. S. Knol: Kennzeichnung der Rauscheigenschaften von Vierpolen (Nachr. Techn. Fachber. 37-40, 1955). (Characterization of the noise properties of fourpoles; in German).

See abstract 2298 and abstract R 277.

**2317:** A. N. Diddens, F. L. H. M. Stumpers and J. Volger: Das Rauschen eines keramischen Halbleiters, dessen Widerstand von der Frequenz und von Magnetfeld abhängig ist (Nachr. Techn. Fachber. 88-89, 1955). (The noise of ceramic semiconductors whose resistance is dependent on frequency and on magnetic field; in German).

The noise voltage of a specimen of the material  $\text{La}_{0.9}\text{Sr}_{0.1}\text{MnO}_3$ , previously shown to have a frequency and magnetic-field dependent resistance, has been investigated. As expected the thermal noise satisfies the Nyquist formula  $\overline{\Delta V^2}/\Delta f = 4kTR(f,H)$ , where  $R(f,H)$  is the real part of the impedance of the specimen at the frequency  $f$  and in the magnetic field  $H$ . A direct current  $i$  increases the noise strongly as in many polycrystalline materials. The spectrum of the current noise is given by  $\overline{\Delta V^2}/\Delta f = K(f,H) i^\alpha f^{-\beta}$  where  $\alpha = 1.3-1.7$ ,  $\beta = 0.9$  and  $K$  is proportional to the value of  $R(f,H)$ .

**2318:** P. Schagen: The "scenioscope", a new television camera tube (T. Ned. Radiogenootschap 20, 291-305, 1955).

Description of a new television camera tube based on the use of a target of semi-conducting glass. See Philips tech. Rev. 17, 189-198, 1955/1956.

**2319:** J. Smit: The spontaneous Hall effect in ferromagnetics, I (Physica 21, 877-887, 1955).

Apart from the normal Hall voltage a magnetized ferromagnetic material usually shows a relatively large extra voltage in the same direction, which can

be found by linear extrapolation to  $B = 0$ . It is shown that this spontaneous Hall effect cannot exist in a perfectly periodic lattice. Measurements at different temperatures suggest that the effect is closely related with the electrical resistivity of the material. Existing theories on the origin of the effect are shown to be invalid, and it is shown that the explanation has to be based on the anisotropic scattering, caused by spin-orbit interaction of the conducting electrons against the imperfections of the lattice.

**2320:** R. Thoraeus, W. J. Oosterkamp, J. Proper, R. Jaeger, B. Rajewsky, E. Bunde, M. Dornreich, D. Lang and A. Sewkor: Vergleichsmessungen der internationalen "röntgen" im Bereich von 8 kV bis 170 kV Erzeugungsspannung (Strahlentherapie, 18 (2), 1955). (Comparative measurements of the international röntgen for generating voltages from 8 kV to 170 kV; in German).

Joint publication from the Radiofysika Institutionen. Stockholm, the Philips Research Laboratories, Eindhoven, the Physikalisch-technische Bundesanstalt, Braunschweig and the Max Planck-Institut für Biophysik, Frankfurt/Main. The main results are given of a comparison between the standard instruments for measuring hard and soft X-rays of the Max Planck Institut für Biophysik and Swedish and Dutch instruments. The measurements with hard X-rays are in good agreement with each other. The comparisons in the soft X-ray region, where errors are by nature larger, show fairly good agreement; studies are planned to improve measuring instruments and methods in this region in order to reduce the still existing differences.

**2321:** J. W. L. Köhler and J. van der Ster: A small liquid nitrogen plant using a gas refrigerating machine (Proc. 9th Int. Congress Refrigeration, Paris, Sept. 1955).

Description of an installation producing liquid nitrogen on a small scale. A rectifying column is coupled to a gas refrigerating machine supplying the necessary cold. The air to be fractionated is not compressed at all, which makes the installation extremely simple. The single column is of the packed type; the incoming air is utilized to reboil the liquid with the higher boiling point. The vapor, issuing at the top of the column, is condensed by the gas refrigerating machine: the resulting liquid is partly utilised as reflux in the column, the rest forming the product. The completely automatic plant is run fully unattended. Relevant constructional details and results are given.

# Philips Technical Review

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

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

## A MAGNETODYNAMIC GRAMOPHONE PICK-UP

### I. CONSTRUCTION.

by N. WITTENBERG:

681.84.081.48:621.395.623

---

*Professional gramophones are commonly equipped with pick-ups of the electrodynamic type (moving coil and fixed magnet). A magnetodynamic type of pick-up will shortly be marketed in which the coils remain fixed while the magnet moves. The new pick-up has various advantages compared with the electrodynamic type, greater sensitivity and lower price amongst them, which make the new pick-up suitable for non-professional equipment. The article below deals the construction of the magnetodynamic pick-up. Its frequency characteristic and some other properties will be treated in a further article.*

---

Of the various types of gramophone pick-up in use the most common is the piezo-electric type<sup>1)</sup>. For professional purposes (in broadcasting studios, and for testing purposes in the manufacture of gramophone records) use is often made of the electrodynamic system: in this, a small coil located between the poles of a permanent magnet follows the movement of the gramophone needle, the signal voltage being induced in it as a result. The construction of the system is in principle the same as that of a moving-coil meter, the gramophone needle taking the place of the pointer moving across the scale.

There are disadvantages inherent in the electrodynamic system. They are manifested particularly when it is desired to employ the system in pick-ups for non-professional equipment of which a high fidelity is required. In such cases the coil must be of extremely light construction, in order to keep the mass of the moving system of the pick-up down to a minimum; otherwise, the resulting accelerations (which may be hundreds of times greater than the acceleration  $g$  due to gravity<sup>2)</sup>) will give rise to inertia forces that will cause severe wear of the gramophone record. To keep wear within acceptable bounds, the mass of the moving system must not be

more than a few milligrams. This implies a coil of extremely thin wire with only a few turns. Consequently the voltage induced in the pick-up is only a fraction of a millivolt. In order to get a good signal-to-noise ratio at the output of the amplifier in spite of this, the amplifier must be provided with a step-up input transformer which has been carefully screened against external magnetic fields (e.g. from the gramophone motor or the amplifier mains transformer). This all adds to the cost.

In order to increase the voltage induced in the moving coil the magnetic field in which the coil moves should be as strong as possible: a rather large permanent magnet is therefore required. The magnet is necessarily situated immediately above the record and hence above the turntable on which the record lies. The turntables of non-professional gramophones are frequently of iron, this being a cheap material and at the same time a heavy one, giving a fly-wheel action. The magnet exercises an attraction on the iron turntable, increasing the needle pressure on the disc. (The weight of the pick-up may be balanced by a spring or counterpoise.) Often the turntables of non-professional gramophones have a diameter of less than 12"; during the playing of the outer part of a 12" disc, the pick-up is not directly over the turntable and the needle pressure is lower than it is further inwards, where the pick-up is directly above. The variation in the vertical needle

<sup>1)</sup> See for example L. Alons, New developments in the gramophone world, Philips tech. Rev. 13, 134-144, 1951/52.

<sup>2)</sup> J. L. Ooms, Philips tech. Rev. 17, 101-109, 1955/56.

force sometimes amounts to several grams in ordinary electrodynamic pick-ups. The increase in the force exerted by the needle results in heavier wear of both needle and disc, and causes losses in the reproduction of high frequencies.

The disadvantages of the electrodynamic system — the need for an input transformer and the unnecessarily high needle pressure on part of the gramophone record — are entirely absent in the magnetodynamic pick-up, which will now be described.

### Design of the magnetodynamic pick-up

In the magnetodynamic pick-up, the magnet and coil exchange their rôles; the magnet moves while the coil is fixed to the body of the pick-up.

A diagram illustrating the principle is given in *fig. 1*. A rod-shaped permanent magnet *M* is located in the air-gap of a yoke *J* of magnetic material, on which the coils *S* are wound. The rod is magnetized in the direction of the arrow, perpendicular to a plane through its axis, and can turn about that axis, being held in two bushes *P* and *R*. At its lower end the rod has an arm *L* (the needle arm) fixed to it, and this arm carries the needle *N*. The lateral movement of

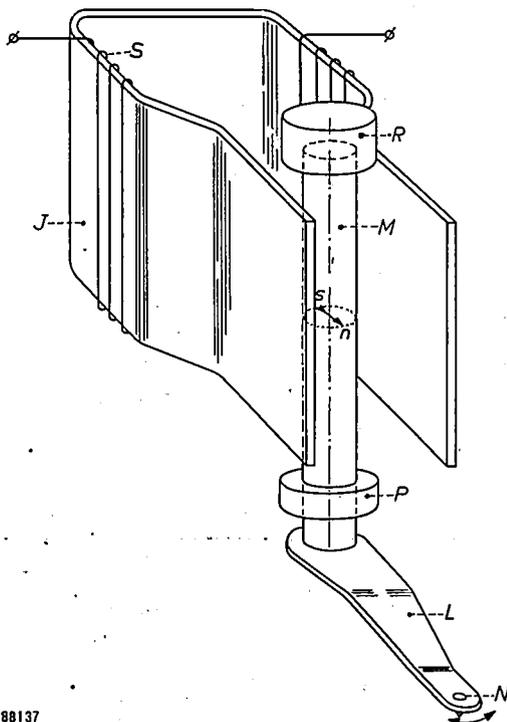


Fig. 1. The essentials of the magnetodynamic pick-up. *M*: ferroxdure rod, magnetised in the direction of the arrow *s-n*. The rod magnet is located between the ends of the yoke *J* carrying the coils *S*, by means of the rubber bush *R* and the polyvinyl chloride bearing *P*, which allow it an angular degree of freedom. The needle arm *L* converts the lateral movement of the needle point *N* into an angular displacement of the magnet about its axis.

the needle point as it follows the modulated groove in the disc, causes the magnet to turn about its axis.

The upper bush *R* is made of rubber. The magnet fits into it tightly and thus suffers a restoring couple when displaced that gives it a definite position of equilibrium. The equilibrium position is made to coincide as far as possible with the position of magnetic symmetry (*fig. 2a*), in which the lines of

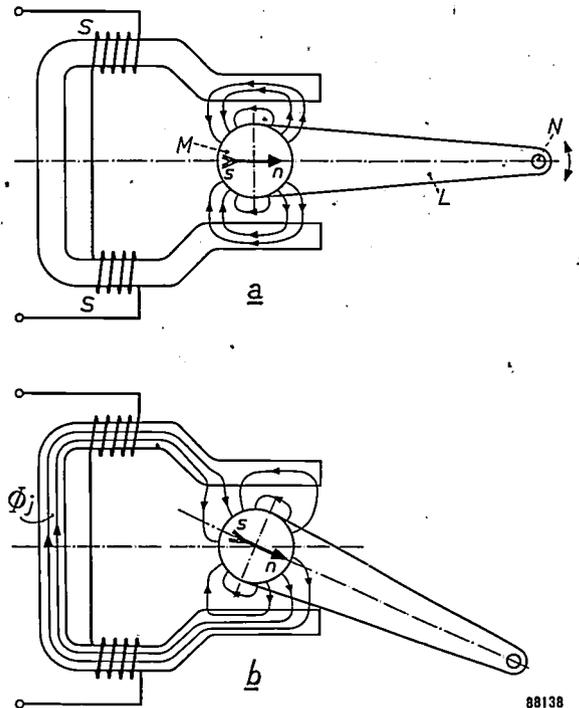


Fig. 2. The magnetodynamic pick-up seen from above (not to scale). (a) Magnet in position of rest; no flux passes through the coils. (b) Magnet turned out of position of rest; a flux  $\Phi_j$  passes through both coils. The letters have the same meaning as in *fig. 1*.

force form a symmetrical pattern, some passing through the air-gap and the ends of the yoke, others through the air only. If the magnet is turned from the equilibrium position the symmetry of *fig. 2a* is upset, and a portion of the flux passes right round the yoke to complete its circuit (*fig. 2b*). The direction of the flux passing round the yoke is dependent on the direction in which the magnet is turned. Lateral movement of the needle point thus produces alternations of flux through the yoke and these induce the signal voltage in the coils.

Two important properties of the above arrangement may be deduced from the following experiment. The magnet, driven by a small motor, is made to rotate between the ends of the yoke, and the voltage set up in the coils is examined. It is found that the voltage is very nearly sinusoidal, this being due to the relatively large air-gaps. From this fact one may

anticipate that the signal voltage induced by the small angular movement of the magnet during actual use will exhibit only extremely slight deviations from linearity (the magnet departs from its position of symmetry by not more than 1° for standard 78 r.p.m. records, and by not more than 0.5° for microgroove records). In the matter of linearity, then, the magnetodynamic pick-up does not fall short of the electrodynamic type, which in theory is strictly linear.

Of the total flux provided by the magnet, a portion  $\Phi_m$  reaches the yoke, while the remainder completes its entire circuit in the air as a stray flux. The second property that may be deduced from the experiment with the rotating magnet is the magnitude of the flux  $\Phi_m$ . For this we require only to know the speed with which the magnet rotates ( $n$  in revs. per sec) and to measure the induced voltage  $E$  (r.m.s. value). The flux is then given by

$$\Phi_m = \frac{E}{4Fnw} \text{ volt.second,}$$

where  $F$  is the shape factor of the voltage function (for a sine wave,  $F = \pi/2\sqrt{2} \approx 1.11$ ) and  $w$  is the total number of turns on the coils. For  $n = 50$  revolutions per second and  $w = 4000$ , we find a voltage  $E = 0.65$  V, from which it follows that  $\Phi_m \approx 0.7 \mu\text{Vsec}$ . This method can be employed in manufacture for testing the magnets.

From the fact that the rotating magnet produces a sinusoidal voltage it follows that  $\Phi_j$ , the flux passing through both coils when the magnet is displaced through an angle  $\alpha$  from the position of symmetry (fig. 3), can be written as  $\Phi_j = \Phi_m \sin \alpha$ . If  $y$  is the displacement of the needle point corresponding to  $\alpha$ , and  $l$  is the distance between the axes of magnet and needle,  $\sin \alpha = y/l$ . Hence

$$\Phi_j = \frac{y}{l} \Phi_m.$$

If the motion of the needle point is sinusoidal, that is, if  $y = \hat{y} \sin \omega t$ , then  $\Phi_j = (\hat{y}/l) \Phi_m \sin \omega t$ . This flux induces an alternating voltage  $e$  in the coils given by

$$e = -w \frac{d\Phi_j}{dt} = -\frac{w \Phi_m \hat{y} \omega}{l} \cos \omega t = -\frac{w \Phi_m \hat{v}}{l} \cos \omega t, \dots \dots (1)$$

where  $\hat{v}$  is the peak value of the velocity of the needle point.

The ratio between the voltage and the velocity is termed the sensitivity of the pick-up; it is usual to take the r.m.s. value of the voltage in millivolts and

the peak value of velocity in cm/sec. Insertion of the values  $w = 4000$ ,  $\Phi_m = 0.7 \times 10^{-6}$  V sec,  $l = 0.5$  cm in equation (1) gives us a sensitivity of about 4 mV per cm/sec. For  $\hat{v} = 5$  cm/sec — a typical

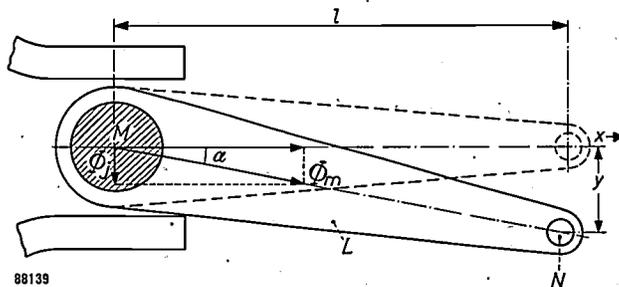


Fig. 3. The same situation as in fig. 2b with the main dimensions drawn to scale.  $\Phi_m$ : the fraction of the total magnetic flux that reaches the yoke.  $\Phi_j$ : the component of  $\Phi_m$  that passes through the coils.

value for microgroove records — this sensitivity gives an r.m.s. voltage of 20 mV. The sensitivity is considerably better than that of electrodynamic pick-ups (without input transformer), but not so good as that of the piezo-electric type. Hence, if the magnetodynamic pick-up is to be connected up to a conventional radio receiver, an extra amplification stage is necessary. A suitable pre-amplifier using a transistor will be discussed in a subsequent article.

**Construction of the pick-up**

The various parts that go to make up the pick-up are shown in fig. 4. We shall now examine these components in some detail and consider the requirements they have to satisfy.

*The magnet*

It follows from equation (1) that, in order to obtain a high voltage from the pick-up (so as to get a high signal-to-noise ratio in the amplifier), the flux  $\Phi_m$  of the magnet must be made as strong as possible. The magnet must further satisfy the following requirements:

1. Its moment of inertia about the axis should be as small as possible in view of the large accelerations occurring.
2. It should be proof against demagnetization by external magnetic fields. One reason for this is that the magnet together with the needle has to be replaced when the needle is worn out; it must not be possible for the loose magnets to become demagnetized by contact with iron parts or tools or by the action of stray fields from transformers, etc.

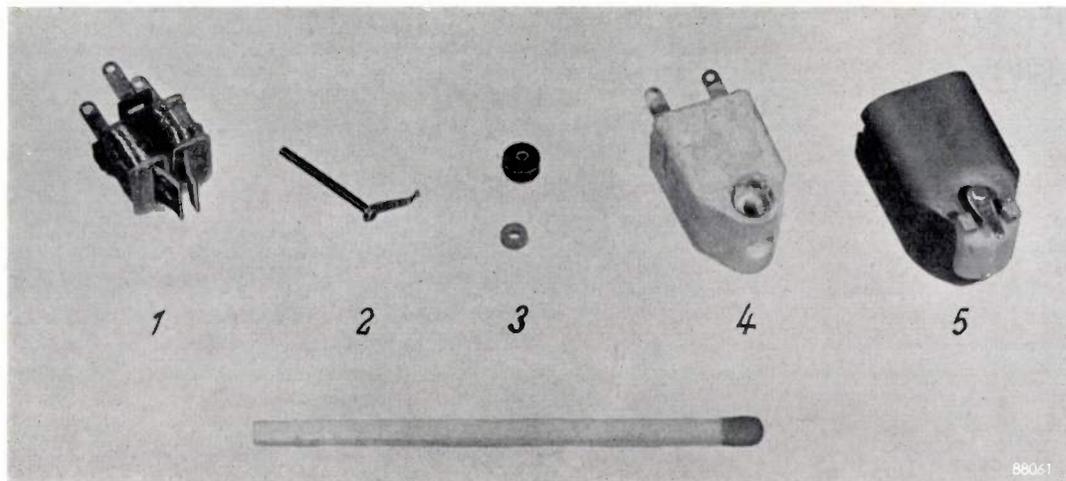


Fig. 4. Component parts of the magnetodynamic pick-up. 1 yoke with coils; 2 magnet with needle arm and needle; 3 rubber bush and p.v.c. bearing; 4 block of polyester resin which is cast around the yoke and coils, place being left for the insertion of magnet and bearings; 5 the block 4 complete with bush, bearing and magnet and shrouded in metal housing for magnetic and electrostatic screening; on either side of the needle arm are protective shoulders.

### 3. The magnet must be easy and cheap to manufacture.

In order to give the magnet a small moment of inertia an unusual design was adopted: a thin rod magnetized perpendicular to its axis. A "short" magnet such as this has a high demagnetization factor and this design was therefore made feasible only by using a material that has a high coercive force. Ferroxdure<sup>3)</sup> is such a material. Its coercive force is about 80 kA/m (about 1000 oersted), in comparison with about 50 kA/m for "Ticonal" steel<sup>4)</sup>. Ferroxdure is suitable for the purpose in mind in other respects as well. Its density is about 4000 kg/m<sup>3</sup>, as against about 7000 kg/m<sup>3</sup> for "Ticonal" (a low density is desirable from the point of view of a small moment of inertia). Thin rods of ferroxdure can be manufactured by extrusion, followed by sintering. These processes lend themselves better to mass production than the casting of rods from magnet steel. After sintering, the ferroxdure rods are ground in a centreless-grinder, since they inevitably become somewhat distorted in the course of the sintering operation. Accurately dimensioned rods are thus obtained which are easily interchangeable and have a restoring couple which is satisfactorily reproducible.

A compromise necessarily has to be made between giving the magnet small mass and making it supply a big flux. The best form has been found to be a rod 1 mm in diameter and 12 mm long, its effective

length being 8 mm. The flux has the value derived above, namely 0.7  $\mu$ Vsec; the mass is about 40 mg, and the effective mass at the needle point, which determines the magnitude of the inertia forces set up, is only 3 mg. (The effective mass is the equivalent mass considered to be concentrated in the needle point and possessing a moment of inertia with respect to the magnet axis equal to that of the whole moving system.)

The attraction between an iron turntable and a magnet of this small size is quite negligible compared with the minimum force with which the needle must rest on the record in order not to jump the groove. This constitutes one of the great advantages of the magnetodynamic system over the electrodynamic system.

#### *The needle arm and the needle*

The function of the needle arm is to communicate to the magnet the movements of the needle point. There are two components of the latter: a lateral movement corresponding to the modulation of the groove, and a vertical movement resulting from the so-called "pinch-effect" (a sinusoidal groove, for example, is narrower in the flanks than at the peaks, with the result that the needle is borne up in the former and sinks down in the latter; see article cited in footnote<sup>1)</sup>). Only the lateral movement of the needle has to be communicated (without distortion or loss) as an angular movement of the magnet; this implies that the needle arm must be so rigid in a lateral sense that resonance of the needle arm and magnet in the lateral direction occurs only at a frequency higher than the range that has to be

<sup>3)</sup> J. J. Went, G. W. Rathenau, E. W. Gorter and G. W. van Oosterhout, Philips techn. Rev. 13, 194-208, 1951/52.

<sup>4)</sup> B. Jonas and H. J. Meerkamp van Embden, Philips tech. Rev. 6, 8-11, 1941.

reproduced, i.e. above 20 kc/s. At frequencies above resonance almost all the movement of the needle is absorbed by the flexion of the needle arm, resulting in a big drop in the output voltage. In the second part of this article, dealing with the frequency characteristic, it will be shown that the needle arm does in fact satisfy the above condition.

Since the needle extends below the needle arm, the lateral force on the needle point creates a couple that to some extent twists the needle arm. Like the frequency of lateral resonance, the resonant frequency of this torsional effect must lie above the range of frequencies to be reproduced. If the system is excited at the torsional resonance frequency, the magnet remains practically still and the output voltage drops to zero, while the needle makes a shrill sound and there is heavy wear of the record. The frequency at which torsional resonance occurs is decided by an effective mass (other than that for lateral movement) at the needle point and the torsional stiffness of the needle arm. The obvious way of raising the torsional resonance frequency above 20 kc/s is to make the needle arm thick enough. However, this would bring us into conflict with a third requirement that the needle arm has to satisfy: it must not transmit to the magnet the vertical movement of the needle resulting from pinch effect; this vertical movement occurs at a frequency double the modulation frequency and, if it contributed to the pick-up output, would introduce a second harmonic. In theory, of course, no movement of the magnet other than the turning movement induces a voltage in the coils; but the slight departures from symmetry that are inevitable in mass-production prevent this ideal from being completely realised. Hence the magnetodynamic pick-up, like the electrodynamic one, does have a certain response to vertical needle movement.

The "vertical response" can be reduced to a minimum by making the needle arm relatively flexible in the vertical direction, whereby the resonant frequency of magnet and arm for vertical movement is reduced to below the frequencies where pinch effect is considerable (roughly speaking above 1000 c/s); hence the vertical movement is absorbed by the needle arm and thus prevented from affecting the magnet. Flexibility of the needle arm in the vertical direction has two further advantages. Firstly, it means that the effective axis for movement in the vertical plane is closer to the needle, so that the effective mass for movement in this direction is virtually limited to the mass of the needle and its fixture to the arm. Secondly, if by mischance the pick-up should be dropped on to the disc, the

needle arm can flex back such that the weight of head is taken on two shoulders fixed to the housing (visible in fig. 4, No. 5); this protects the brittle ferroxdure rod against breakage. On the other hand, however, the needle arm must be sufficiently stiff to support the weight which the pick-up exerts, via the needle, on the record (about 10 grams) without bending too far. By suitable choice of the material, shape and dimensions of the needle arm these requirements can be satisfied. The resulting torsional resonance frequency is in the region of 25 kc/s while the resonant frequency for lateral vibration is still higher.

From what has been said it will be evident that the needle arm, apparently so simple, is a component to which a great deal of attention has to be given.

The magnet and the needle are fixed to the needle arm by cleating them with small collars of aluminium. Needle, collar and needle arm are first positioned in the cleating jig. The unflanged end of the collar is then cleated by a suitably shaped tool (see fig. 5) The magnet is fixed to the needle arm by

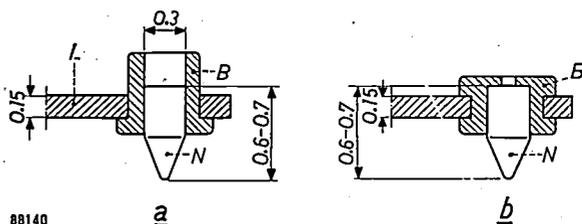


Fig. 5. Attachment of the needle *N* to the needle arm *L* by an aluminium cleat *B*. (a) Before cleating, and (b) after. Dimensions are in mm.

a similar method. Not only does this method give a reliable joint but all play is taken up by the aluminium cleating so that there is no need to demand extremely fine tolerances in the dimensions of the parts.

The magnetodynamic pick-up (fig. 6) is manufactured in two models, identified by a coloured mark, which differ only in the type of needle and which are easily interchangeable. Type AG 3020 (with a green spot) is for standard 78 r.p.m. records and has a sapphire needle; the radius of curvature of its tip is 75 microns, as is usual in needles for these records. Type AG 3021 (red spot) is for microgroove records; the tip of its diamond needle has a radius of curvature of 25 microns.

Diamond was chosen for the needle point of the second type for the following reasons. With the force of 10 g weight with which the point rests on the record and the smaller radius of curvature of the microgroove needle, the pressure is so high that a sapphire needle wears comparatively quickly. The wear of the needle results in distorted reproduction, particular-

ly of high notes. If the pick-up is used with an amplifier and loudspeakers capable of reproducing very high audio frequencies<sup>5)</sup>, the wearing of the sapphire point is perceptible as a troublesome distortion after comparatively few records have been played. The rate at which diamond wears is some tens of times slower. Added to this, less wear of the needle point means less wear in the records. One might expect that the harder and more resistant to wear the needle point, the greater the wear of the record but in fact the latter is mainly dependent on the shape of the needle point: the harder the needle, therefore, the longer it will retain its initial approximately hemispherical shape and hence the smaller the wear of the record.

must not be greater than necessary; at low frequencies, the stiffness largely determines the force required to give the needle a certain lateral displacement, and this force in turn determines the minimum downward force of the needle required to prevent it jumping the groove. We shall return to this matter of choosing a value for the vertical force later.

A value of 200 N/m ( $= 0.2 \times 10^6$  dyne/cm) has been chosen for the stiffness  $s$ . The maximum amplitude  $\hat{y}_{\max}$  is 100 microns for a standard



Fig. 6. Philips record-changer, type AG 1003, equipped with the AG 3021 magnetodynamic pick-up.

### The magnet bushes

The magnet bushes must allow the magnet to turn about its axis, but provision also has to be made for a restoring couple that tends to bring it back to a definite position of equilibrium. This is necessary to ensure that the needle arm lies parallel to the groove when the latter is unmodulated. If this condition is satisfied, tracing distortion is at a minimum<sup>6)</sup>.

The restoring couple is obtained by clamping one end of the magnet into a rubber bush ( $R$  in fig. 1). The stiffness so obtained must have a certain minimum value for the reason given above, but it

78 r.p.m. record at low frequencies; at higher frequencies the maximum amplitude is less, being limited by other factors during the cutting of the record: see article quoted in footnote<sup>2)</sup>. The lateral force necessary to give the needle point this maximum amplitude is therefore of the order of  $s\hat{y}_{\max} = 200 \times 100 \times 10^{-6} \text{ N} = 20 \text{ mN}$ .

Apart from the restoring couple, the moving system must have a definite amount of damping. This is a question not only of the resonance phenomena already mentioned but also of the resonance of the mass of the pick-up arm with the stiffness of the moving system. In Philips record-changers equipped with a magnetodynamic pick-up this latter resonance occurs at about 30 c/s. The damping action of the rubber bush is very slight; for this reason, at the lower end a polyvinyl chloride bush ( $P$  in fig. 1) is used in which the magnet can turn. The plastic gives considerable mechanical damping

<sup>5)</sup> See for example J. J. Schurink, The twin-cone moving-coil loudspeaker, Philips tech. Rev. 16, 241-249, 1954/55.

<sup>6)</sup> To discuss the reasons for this would be outside the scope of the present article. It may be mentioned here that the arm of the pick-up should ideally be designed so that the symmetry plane of the tracking system is always parallel the tangent to the unmodulated groove (see for example to B. B. Bauer, Tracking angle in phonograph pick-ups, Electronics 18, 110-115, March 1954).

over an extensive range of frequency, but the restoring couple set up by the lower bush is very small. Together with the effective mass and the stiffness, the damping provided by the lower bush goes to make up the mechanical impedance; it must not be made excessive, because a damping force is also supplied by the needle point. This being so, the magnet is not clamped in the polyvinyl chloride bush but can rotate as in a bearing; in fact the bore of the bush is somewhat greater than the diameter of the magnet so that the latter is pulled to one side of the bearing wall by the frictional force between needle and record. This turns out to give a satisfactory degree of damping, the fact that the bush in question is lower down and very close to the needle also playing a part. Interchanging the two bushes has little effect on the restoring couple but causes a big decrease in damping.

The mounting of the magnet described above is not rigid in any direction and hence it allows of movements other than rotation. This is undesirable, as we have already seen, in view of the second harmonic that may arise from the pinch effect. On the other hand, mounting the magnet in bearings that permit a turning movement only appears to make reproduction "hoarse", particularly at high frequencies. There is a rule valid for all pick-up systems according to which the moving system of the pick-up must be given various degrees of freedom in the higher frequency range; the limited movements thereby permitted keep the mechanical impedance to a minimum. As far as rotational movements are concerned, the impedance is a minimum for rotation about that axis for which the moment of inertia is least. At high frequencies the horizontal and vertical movements of the needle point, and hence those of the whole moving system, are very complicated and very difficult to investigate; their amplitudes amount to a few microns at the most and their accelerations attain some hundreds of times  $g$ , as already stated. If the moving system is deprived of one degree of freedom, the mechanical impedance at the needle point may rise so high that the forces set up by the groove modulation acquire values at which a distortionless trace is no longer possible. In every pick-up, therefore, a favourable compromise must be sought between the number of degrees of freedom allowed to the moving system and the acceptance of voltages produced by undesired movements of that system. The quality of the pick-up depends considerably on how well this compromise is chosen. In this respect a system such as the magnetodynamic, in which in principle only one type of movement induces a voltage, has a great

advantage over a system wherein voltages are induced by more than one type of movement.

#### The yoke

The yoke must have as small a reluctance as possible, and is therefore made of a material of high initial permeability. It consists of a strip of cross-section of  $6 \times 0.35 = 2.1 \text{ mm}^2$ , bent into the required shape.

If the needle point has an amplitude  $\hat{y}$  of 0.1 mm, the amplitude of the alternating flux through the yoke is:

$$\Phi_j = \Phi_m \frac{\hat{y}}{l} = 0.7 \times 10^{-6} \times \frac{0.1}{5} = 14 \times 10^{-9} \text{ volt.sec.}$$

Hence the magnetic induction in the yoke in the same circumstances is:

$$B_j = \frac{14 \times 10^{-9}}{2.1 \times 10^{-6}} \approx 7 \times 10^{-3} \text{ volt.sec/m}^2 (= 70 \text{ gauss}).$$

This is a very low figure, and hence the possibility of distortion due to the saturation of the yoke is ruled out (the saturation induction for the material used is about 0.5 Vsec/m<sup>2</sup>). Equally, there is little likelihood that magnetic asymmetry will produce too high an induction in the yoke, for even if the entire flux  $\Phi$  passed through it, the static induction in the yoke would still only amount to 0.35 Vsec/m<sup>2</sup>.

The low value of the alternating induction has an advantage over and above the elimination of distortion due to saturation. There is no danger of noise arising from the reorientation of Weiss domains (Barkhausen effect<sup>7</sup>). The absence of this kind of noise has been confirmed by an experiment in which the needle was made to move with a large amplitude at a low frequency; there was absolutely no indication of any noise associated with this movement.

Polyester resin is cast around the yoke carrying the coils. A sturdy block (fig. 4, No. 4) is thus formed, in which the positions for the bearings are accurately centred with respect to the yoke. The coils are completely enclosed, making the pick-up proof against tropical climates.

The two coils together constitute an approximately astatic system and hence they are but little affected by stray magnetic fields arising from the gramophone motor and any transformers that may be in the vicinity. On account of the relatively high output voltage, the leads from the pick-up to the amplifier pick up little interference. In both these respects the magnetodynamic pick-up compares

<sup>7</sup> H. Barkhausen, Phys. Z. 20, 401-403, 1919; B. van der Pol, Versl. Kon. Akad. Wet. Amsterdam, 29-I, 341-348, 1920.

favourably with electrodynamic and other magnetic types. Nevertheless, as a further measure against any residual fields, it is provided with magnetic screening. This consists of a soft iron casing (fig. 4, No. 5) into which the polyester block is fixed once the two bearings have been inserted in it.

Finally, the system thus screened is mounted in a "Philite" housing. The housing has a terminal by which the magnetic screening can be connected to the earth terminal of the amplifier, so that it serves as electrostatic screening at the same time. The yoke is also connected to earth, thus obviating undesirable phenomena due to static electricity generated by the friction between needle and record.

#### The optimum value for the vertical needle force

$F_v$ , the vertical force with which the needle presses on the record, is an important quantity. In general, it is desirable to keep this force small to prevent

both cases  $F_{n1}$  is balanced by a reaction from the left-hand wall on the needle point at  $Q_1$ , their point of contact. In fig. 7a the direction of  $F_{n2}$  is towards the other wall and it is balanced by the reaction at point  $Q_2$ , the needle thus being kept in contact with the wall at this point. However, in fig. 7b the direction of  $F_{n2}$  is away from  $Q_2$ ; the needle will therefore tend to lose contact with the right-hand wall and, if  $F_{n2}$  is sufficient to overcome the frictional force due to movement against the left-hand wall, the needle will ride up the latter. This results in distortion or, at worst, de-tracking of the needle. Hence the highest value of  $F_l$  that occurs determines the lowest permissible value of  $F_v$  if this danger be avoided.

As already stated,  $F_l$  is the resultant of a number of lateral forces. Of these, the inertial force of the moving system (its magnitude is the product of the effective mass at the needle point and its acceleration) is the largest at high frequencies. In order to minimise

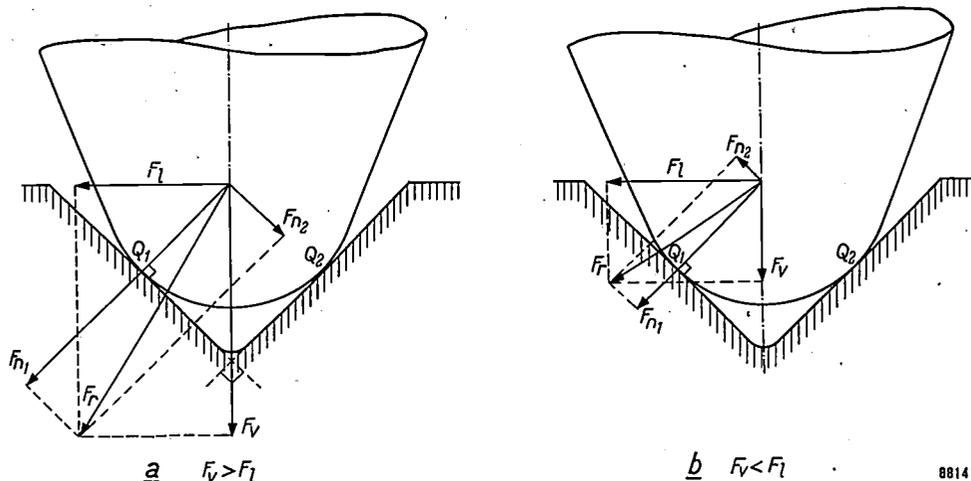


Fig. 7. Forces on the needle point.  $F_v$  vertical force with which the needle rests on the disc.  $F_l$  total lateral force. In case (a)  $F_v > F_l$ , in case (b)  $F_v < F_l$ .  $F_r$ , the resultant of  $F_v$  and  $F_l$ , has components  $F_{n1}$  and  $F_{n2}$  perpendicular to the two walls of the groove. In case (a)  $F_{n2}$  is in a direction such that the needle maintains contact with the right-hand wall at point  $Q_2$ ; in case (b) the direction of  $F_{n2}$  is such that this contact is not maintained and the needle rides up along the left-hand wall of the groove. (Friction, and the deformation of the walls under needle pressure, are neglected.)

88141

excessive wear. If the pick-up is too light, on the other hand, there is a risk that the needle will jump the groove. The following considerations show us in what circumstances this risk is present.

Apart from the vertical force  $F_v$ , there is a lateral force  $F_l$  acting on the needle point, which is the resultant of several lateral forces. Diagrams illustrating two cases appear in fig. 7, (a) where  $F_v > F_l$  and (b) where  $F_v < F_l$ .  $F_r$ , the resultant of  $F_v$  and  $F_l$ , can be resolved into two components,  $F_{n1}$  and  $F_{n2}$ , perpendicular to the walls of the groove, which are approximately at right angles to each other. In

this force the moving system is designed to have as small an effective mass as possible; the value of the latter is in fact 3 mg.

The maximum permissible acceleration is found by considering the smallest radius of curvature that may occur in the groove without distortion arising; if the radius of curvature  $\rho$  of the modulation peaks were smaller than  $R$ , the radius of curvature of the needle point, the needle would be in contact with the groove at three points instead of two, and distortion known as over-modulation would arise. Taking  $\rho = R$  as a rough basis, a peak acceleration

of the order of 300 g is found for the fundamentals of the higher frequencies (see article cited in footnote<sup>2</sup>). However a third harmonic of about 40% is then present in the velocity<sup>8</sup>), and consequently one of  $3 \times 40 = 120\%$  in the acceleration. Taking this into consideration, we arrive at a maximum value of the total acceleration of  $(100 + 120)\%$  of 300 g, that is, of 660 g  $\approx 6600$  m/sec<sup>2</sup>. This, with the effective mass of 3 mg, produces a maximum lateral force of inertia of  $3 \times 10^{-6} \times 6600$  N  $\approx 20$  mN ( $\approx 2$  g weight).

Oscillograph measurements of the voltage delivered by the pick-up shows that accelerations even greater than 6600 m/sec<sup>2</sup> — and hence correspondingly greater forces of inertia — sometimes occur at the needle point. The higher accelerations may be caused by higher modulation velocities (where  $\rho < R$ ), giving overmodulation, which brings about a big increase in distortion; this in its turn again increases the acceleration owing to the harmonics involved. A further cause of the increased accelerations may be the building up of high amplitude needle vibrations due to groove-needle resonance<sup>9</sup>).

There are two other lateral forces besides the force of inertia just stated. Firstly, owing to the frictional drag of the needle on the record and the geometry of the pick-up arm, there is a constant inward force on the pick-up across the record; in Philips gramophones this is about 15 mN. Secondly there is the stiffness force which was shown earlier to have a maximum value of 20 mN at the lower frequencies for normal 78 r.p.m. records, this being the product of the stiffness  $s = 200$  N/m and maximum needle displacement  $\hat{y}_{\max} = 100$  microns. Such large displacements occur only at low frequencies, but the higher frequencies and their harmonics can also be present in the groove at the same time. These three forces can therefore be additive and in this way we arrive at a maximum of  $20 + 15 + 20 = 55$  mN for the lateral force  $F_l$ ; this, then, is the lowest value permissible for  $F_v$ . There is still a further point to be considered, however: the pinch effect causes a periodic variation

in the magnitude of  $F_v$ . The amplitude of the alternating component may amount to about 10 mN so that a static value of at least  $55 + 10 = 65$  mN is required for  $F_v$ .

This calculation is by its very nature only an approximation to the real state of affairs. However, bearing in mind that all the unfavourable circumstances are rarely present in combination, the minimum value of 65 mN, as calculated, agrees well with the value found by experiment. In fact a force of from 60 to 70 mN appears to be just sufficient for records with large modulation amplitudes to be faultlessly played, provided that the greatest attention is paid to the bearings and the balancing of the pick-up arm. Extreme care in production is not an insuperable objection in the making of professional equipment but, in the case of gramophones for domestic use, it is required that efficient tracing of the groove be obtained with pick-up arm that is simple and inexpensive. For these reasons  $F_v$  is given a somewhat higher value, viz. 100 mN (about 10 g weight). This value not only guarantees stable operation but also minimum wear, for it has been shown experimentally that with a weaker force there is again an increase in wear.

Part II of this article will deal with some properties of the magnetodynamic pick-up, including its frequency characteristic.

**Summary.** The new magnetodynamic pick-up type AG 3020/21 has a small rod-shaped magnet as its moving system; the rod is magnetized perpendicularly to its axis, about which it can turn, and is mounted between the ends of a yoke of magnetically soft material. A needle arm is fixed to the rod magnet whereby the lateral movement of the needle as it follows the groove in the record is converted into an angular movement of the magnet. An alternating flux is thus produced in the yoke, giving rise to a signal voltage in the coils wound on it. The angular movement of the magnet is provided for by an upper flexible bush of rubber and a bearing at the lower end of polyvinyl chloride. The rubber bush gives the magnet a restoring couple and the p.v.c. bearing provides the necessary damping against undesired resonances. The pick-up is manufactured in two models, type AG 3020 with a sapphire needle fitting the groove in standard 78 r.p.m. discs, and type AG 3021 with a diamond needle for microgroove discs. The departure from linearity between the angular movement of the magnet and the induced signal voltage is extremely slight. The sensitivity of the pick-up (ratio of r.m.s. voltage to peak needle velocity) is about 4 mV per cm/sec. A study of  $F_v$ , the force with which the needle presses on the disc, yields an optimum value of  $F_v$  of 100 mN ( $\approx 10$  g weight) in gramophones for domestic use, this value being sufficient to prevent de-tracking and giving minimum wear. The pick-up is proof against tropical climatic conditions.

<sup>8</sup>) J. A. Pierce and F. V. Hunt, On distortion of sound reproduction from phonograph records, *J. Acoust. Soc. Amer.* **10**, 14-28, 1938/39; W. D. Lewis and F. V. Hunt, A theory of tracing in sound reproduction from phonograph records, *J. Acoust. Soc. Amer.* **12**, 348-365, 1940/41.

<sup>9</sup>) J. B. S. M. Kerstens, Mechanical phenomena in high-note reproduction by gramophone pick-ups, *Philips tech. Rev.* **18**, 89-97, 1956/57, in particular pp. 000. See also Part II of present article.

## THE NOISE EMISSION OF BALLASTS FOR FLUORESCENT LAMPS

by E. W. VAN HEUVEN.

621.327.534.15.032.434:  
621.318.424:534.833

*The introduction of fluorescent lighting into offices, shops and the home, required a reduction of the noise emission to a level compatible with the "silence" desirable in such surroundings. This article describes how the noise problem was solved, by special methods of construction and mounting and by acoustic insulation. Both the theoretical basis of these solutions and the associated measurements are discussed.*

Unlike incandescent lamps, gas-discharge lamps, such as tubular fluorescent lamps ("TL" lamps), cannot be directly connected to the mains but must be put in series with some type of ballast to limit the current in the discharge tube.<sup>1)</sup>

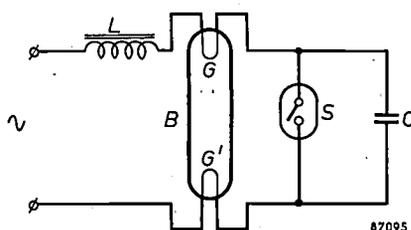


Fig. 1. Diagram of a fluorescent lamp with ballast. *B* lamp with filaments *G*, *G'*, *L* choke, *S* starter with interference-suppression capacitor *C*.

A frequently-used circuit, in which the ballast consists of a self-inductance (choke), is shown diagrammatically in *fig. 1*. When the lamp is switched on, the two filaments *G* and *G'* of the discharge tube are connected in series with each other and with the choke via the starter *S*. The discharge circuit is then short-circuited and a fairly large current passes through the filaments, which quickly heat up. When the connection is broken at *S* (this happens automatically), the discharge is initiated. After each half phase of the alternating current the lamp extinguishes and is re-ignited, now, however, without the intervention of the starter since, on account of their thermal inertia, the filaments are still sufficiently hot, and moreover, there are sufficient residual ions and electrons in the gas to initiate the discharge at a voltage only slightly higher than the working voltage.

With the increasing use of fluorescent lamps for indoor lighting (offices, living-rooms) more attention came to be paid to the noise associated with these

installations. The noise, which includes both low notes (hum) and high notes (rustle), originates principally in the ballasts. The cause lies in the alternating magnetization and the consequent deformation of the iron core of the ballast.

We shall first discuss the nature of the noise and the manner in which vibrations of the ballast are communicated to the air. We shall then go on to deal with the theoretical aspects of core deformations, confining ourselves for the time being to low frequencies. From these considerations, the precautions that can be taken to combat ballast hum are derived. Lastly, we shall deal more fully with the causes of rustle and how it can be combated.<sup>2)</sup>

### The nature of the noise

*Fig. 2* shows oscillograms recorded on a fluorescent lamp with a choke; *I* is the voltage across the lamp, *II* the voltage across the choke, *III* the current flowing through the lamp and consequently through the choke. To a first approximation, this current is sinusoidal. Closer examination shows that there are certain deviations from the sinusoidal form, so that in addition to the mains frequency (50 c/s), higher frequencies occur, which are odd multiples of the mains frequency.

The changes in the shape of the core of the coil are, as we shall see below, periodic, having a fundamental frequency twice that of the mains (i.e. 100 c/s). Apart from this frequency, "higher harmonics" occur in the vibrations of the core, whose frequencies are small multiples of 100 c/s. These low-frequency vibrations are responsible for *hum*, a noise that is widely encountered with alternating current apparatus (it can be heard, for example, if the ear be placed close to a transformer housing).

<sup>1)</sup> For a description of ballasts, see *Fluorescent Lighting*, edited by C. Zwicker, Philips Technical Library, 1952.

<sup>2)</sup> See also: E. W. Heuven, On the noise of fluorescent lighting installations, *Acustica* 5; 101-111, 1955 (no. 2).

Superimposed upon the approximately sinusoidal current there are other irregularities which take the form of small current peaks of short duration. This may be seen from the rapid voltage variations found

oscillations. Current peaks (a) and (b) are responsible, as will be described in greater detail later, for higher frequency vibrations in the core (1000-3000 c/s), which give rise to the noise known as *rustle*.

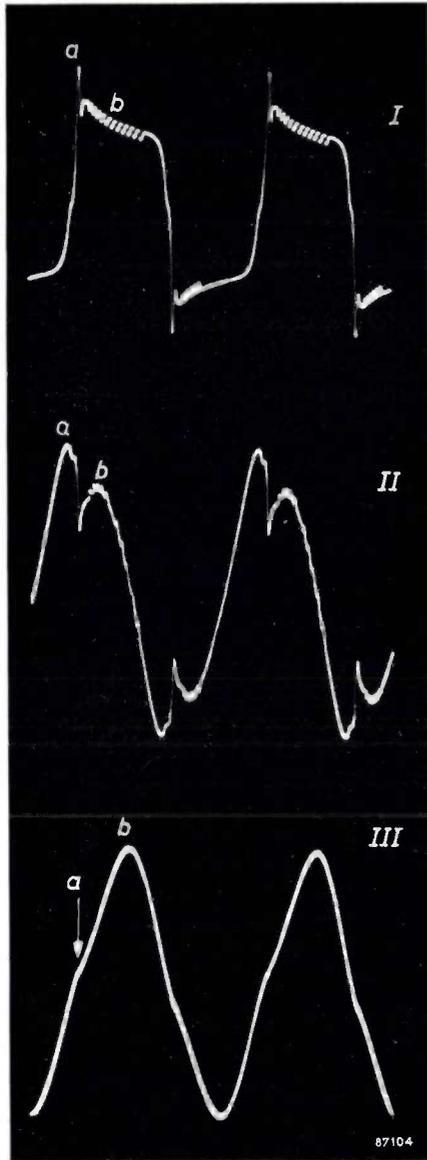


Fig. 2. Oscillograms of the voltage across the lamp (I), the voltage across the choke (II) and the current through the lamp and the choke (III). Re-ignition occurs at *a*, giving rise to voltage peaks. The voltage peaks at *b* result from irregularities in the discharge. Both phenomena (*a* and *b*) give rise to variations in the current (barely discernible in III).

in curves I and II in fig. 2 in two places, viz. *a*) at the moment of re-ignition, *b*) during the time that the current is large. (The corresponding current peaks are not sufficiently large to be visible in the oscillogram of the current, curve III, owing to the high impedance of the choke for rapid variations.) Peaks (*b*) can have two causes, viz. irregular discharges in the gas around the filaments, and plasma

In addition to the noises designated hum and rustle, a third troublesome noise sometimes occurs, which might be called "rattle". This is likewise caused by mechanical vibration of the ballast and its fixture, when these contain mutually supporting parts that are held together only by gravity. If, as a result of vibration, vertical accelerations occur somewhere in the construction that are greater than "g" the acceleration due to gravity (for a frequency of 100 c/s, this occurs when the amplitude is greater than 25 microns), these parts move with respect to each other and cause rattle. Since this noise is less important, and can usually be easily rectified, we shall not discuss it further.

#### Transfer of vibrations from the ballast to the air

The way in which the vibrations of the ballast reach the ear is different for hum than for rustle. Sound frequencies of the order of 100 c/s correspond to wavelengths of several metres, i.e. much greater than the dimensions of the ballast. A ballast suspended freely in space (or elastically suspended) would therefore be a very poor radiator of such waves<sup>3)</sup>, and would give very little hum. It is more customary, however, to attach the ballast firmly to a wall, to the ceiling or to a fixture. The surfaces of the latter objects, which are brought into sympathetic vibration with the ballast, are not small compared with the wavelengths mentioned, so that they will emit a perceptible hum. The same phenomenon is involved here as in the well-known experiment with a tuning fork. The fork vibrating freely in the air is scarcely audible, but is clearly audible if its stem is held on a table top or similar surface.

With rustle the frequencies involved are of the order of 3000 c/s, corresponding to wavelengths of about 10 cm; the ballast itself is a suitable radiator for such notes, so that rustle will be audible even in the absence of any sounding-board.

To avoid noise from the ballast it is therefore necessary to prevent vibrations from being transmitted to walls, etc. The hum is then inaudible. To combat rustle it is necessary, in addition, to suppress direct emission by the ballast in the frequency range 1000-3000 c/s. We shall now analyze the phenomenon of hum in greater detail.

#### The origin of hum

The deformation of an iron core by magnetization may result from:

<sup>3)</sup> See, for example, A. Th. van Urk and R. Vermeulen, The radiation of sound, Philips tech. Rev. 4, 213-222, 1939.

- a) elastic deformation in iron cores with an (air) gap due to attraction between the parts of the iron circuit on either side of the gap.
- b) magnetostriction.

The term magnetostriction comprises a whole series of deformation phenomena, shown by ferromagnetic materials on magnetization (as well as,

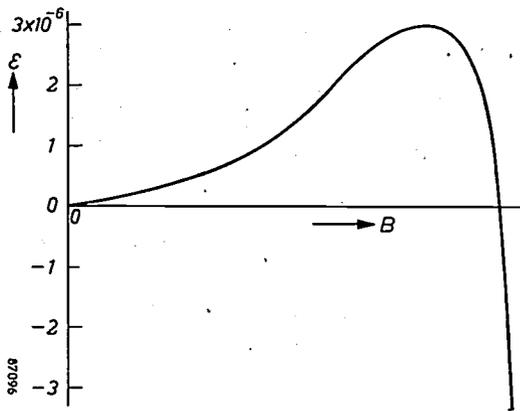


Fig. 3. Magnetostriction of transformer lamination.  $\epsilon = \Delta l/l$  is plotted as a function of the induction  $B$ . As saturation is approached  $\epsilon$  changes sign.

in a number of cases, the reciprocal effects, i.e. change of the magnetic properties by the application of a mechanical load). The most important of these effects is the change in length in the direction of magnetization<sup>4)</sup>. The relative change in length  $\epsilon = \Delta l/l$  is positive for iron at low magnetizations, but changes in sign as the magnetization approaches the saturation value. Fig. 3 gives a general idea of  $\epsilon$  as function of the magnetization  $J$  or, what amounts to practically the same thing, as a function of the induction  $B$ <sup>5)</sup>.

The beginning of the curve is approximately described by the formula:

$$\epsilon = \frac{\alpha}{2\mu_0} B^2, \dots \dots \dots (1)$$

where  $\mu_0 = 4\pi/10^7$  volt seconds/ampere metre, and  $\alpha$  is a constant, which for transformer iron, depending upon the silicon content and the pre-treatment of the iron, amounts to  $1.3 \times 10^{-12}$  metres per newton (m/N). The induction in a choke core is seldom very high in practice, so that formula (1) is usually valid. It may be seen from the formula that the magnetostrictive change in length is approximately sinusoidal and has a frequency double that of the current, which is considered to be sinusoidal.

<sup>4)</sup> W. D. Williams, Magnetostrictive phenomena, General Electric Rev. 45, 161-163, 1942. S. C. Leonard, Magnetostriction made visible, idem 45, 637-641, 1942.

<sup>5)</sup> For other ferromagnetic materials, or for single crystals of iron oriented in a particular way, these curves may be different: see the literature cited in <sup>4)</sup>.

The induction  $B$  then varies<sup>6)</sup> as  $\sin^2 \omega t$ , i.e. as  $\frac{1}{2}(1 - \cos 2\omega t)$ .

If care is taken in designing the iron core to ensure uniform magnetization, then the magnetostrictive expansion leads to an approximately proportional increase in the external dimensions in a plane parallel to that of the laminations<sup>7)</sup>. Since at the corners of the core the induction is somewhat greater on the inside of the bend than on the outside, there occurs at the same time a slight bending, which we shall ignore for the time being.

**Measures to reduce hum**

Having regard to the contribution of elastic deformation (a), we must distinguish between iron cores with air-gaps (for chokes) and cores without air-gaps (for transformers). We shall first examine the simplest case: symmetrical cores with air-gaps.

*Symmetrical cores with air-gaps*

Consider a core of the design shown in fig. 4. The air-gap is located in a plane of symmetry of the core which passes through the centre of gravity. We shall assume for the moment that the ballast is freely suspended in space or is elastically suspended in such a way that it can move freely, the centre of gravity thus remaining stationary.

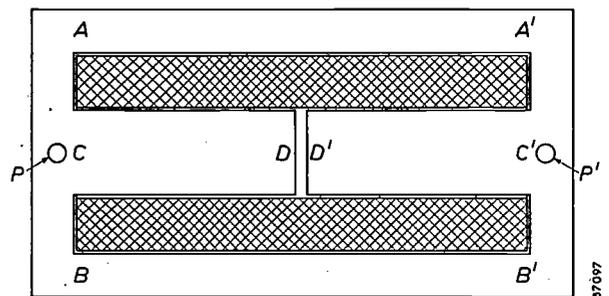


Fig. 4. Cross-section of a choke with a symmetrical iron core. The core consists of three limbs  $AA'$ ,  $BB'$  and  $CC'$  connected by yokes. The air-gap is between  $D$  and  $D'$  in the centre limb and can be filled with non-magnetic material.  $P$  and  $P'$  are the points of attachment.

<sup>6)</sup> If the (odd) higher harmonics are taken into account in the expression for the current ( $i = A \sin \omega t + B \sin 3\omega t + C \sin 5\omega t + \dots$ ), then on squaring such terms as  $\sin 4\omega t$  are found, i.e. frequencies that are multiples of double the mains frequency.

<sup>7)</sup> Together with the elongation by a factor  $(1 + \epsilon)$  in the direction of magnetization, there occurs a contraction by a factor  $(1 - \epsilon/2)$  at right angles to the direction of magnetization (magnetostriction causes hardly any variation in volume). The transverse contraction of the yokes of a core, see fig. 4, thus acts in opposition to the lengthwise expansion of the limbs, and vice versa. Nevertheless positive magnetostriction results in an increase in the periphery of the laminations since the limbs and yokes are appreciably longer than they are broad.

When the iron is magnetized, the poles on either side of the gap attract each other and the two halves  $CD$  and  $C'D'$  of the centre limb are stretched (the other sections of the iron circuit are also deformed: the outer limbs  $AA'$  and  $BB'$  are compressed and the yokes  $AB$  and  $A'B'$  are bent). However, the elastic deformation is counteracted by the magnetostriction, for the latter tends to increase the dimensions of the core in the plane of the laminations (plane of the diagram in fig. 4), that is to say, magnetostrictive deformation implies that points on the line  $CC'$  move away from the centre of gravity, the displacement being proportional to the distance to the centre of the gap. With elastic deformation it is just the reverse: here the displacement is greatest at the gap and less towards the outside. On line  $CC'$  therefore there should be two points — located symmetrically with respect to the centre of gravity — where the total displacement is nil. It has been found that the elastic and the magnetostrictive deformations depend in an identical manner upon the induction (and, therefore, upon the current through the coil), so that during the whole vibration, resulting from variations of the current, both points mentioned remain stationary. In practice the elastic displacement is found to be greater than the magnetostrictive displacement at all points within the core, so that these nodal points lie outside the core.

If the air-gap now be filled with a material with the same magnetic properties as air, but whose rigidity is such that the elastic deformation is reduced, then the vibration amplitudes of all points in the core are reduced, giving a considerable reduction in the hum. Moreover, the two nodal points are brought closer to the centre of gravity, and may even lie within the core. The location of these points is governed by the dimensions and by the filling material which can be so chosen that these points exactly coincide with  $P$  and  $P'$ , the points of attachment of the choke. Vibration cannot then be transmitted to the wall.

The above ideas, when worked out quantitatively, lead to the following formula for the increase in distance between two arbitrary points  $P$  and  $P'$  situated symmetrically at a distance  $L$  from each other on the centre limb:

$$\Delta L = \frac{B^2}{2\mu_0} \left[ aL - \left( \frac{O_1}{S_1 + S_2} \right) \left( 1 - \frac{S_1}{O_1 E_1 / L} \right) \right] \quad (2)$$

where  $O_1$  is the cross-section,  $E_1$  Young's modulus of the material of the core, and  $S_1$  and  $S_2$  are the respective "rigidities" of the core, and the gap-filling material. In this formula  $S_2 = O_2 E_2 / \delta$  should

be regarded as the unknown ( $O_2 =$  cross-section and  $E_2 =$  Young's modulus of the filling material,  $\delta =$  the gap width).  $S_2$  must be given such a value, by a suitable choice of material, that  $\Delta L = 0$ .

In deriving (2) it has been assumed that the elastic and magnetostrictive deformations may be calculated as in the static case. For low hum frequencies this is justified (the forces and the deformations are in phase).

The poles on either side of the air-gap (at  $D$  and  $D'$ ) are drawn together by a force

$$F = \frac{1}{2} BHO_1 = \frac{1}{2\mu_0} B^2 O_1 \quad \dots \dots \dots (3)$$

( $H$  is the magnetic field in the gap).

By filling the gap with an elastic material,  $D$  and  $D'$  are pushed away from each other with a force

$$F_2 = -\frac{O_2 E_2}{\delta} \Delta \delta = -S_2 \Delta \delta \quad \dots \dots \dots (4)$$

Here,  $\Delta \delta$  is the increase in the distance  $\delta$  between  $D$  and  $D'$ .

The resultant force  $F - F_2 = F_1$  gives rise to a complex deformation of the core: limbs  $AA'$  and  $BB'$  are compressed and bent, parts  $AB$  and  $A'B'$  are mainly bent and parts  $CD$  and  $C'D'$  are elongated. These deformations result in a reduction  $-\Delta \delta$  of  $\delta$  at the level of the gap.  $-\Delta \delta$  is proportional to  $F_1$ :

$$F_1 = -S_1 \Delta \delta \quad \dots \dots \dots (5)$$

By analogy with equation (4), the proportionality factor  $S_1$  can be regarded as a rigidity. Owing to the complex character of the deformation corresponding to  $S_1$ , the latter cannot be expressed simply in other quantities. Combination of (3), (4) and (5) gives:

$$\Delta \delta = -\frac{B^2}{2\mu_0} \left( \frac{O_1}{S_1 + S_2} \right) \quad \dots \dots \dots (6)$$

The extension of each of the parts  $CD$  and  $C'D'$  is such, that two points in such a part, originally separated by a distance  $l$ , are at a distance  $l + \Delta l$  from each other, where

$$\Delta l = l \frac{F_1}{O_1 E_1}$$

If we assume that points  $P$  and  $P'$  (separated by a distance  $L$ , of which  $L - \delta$  is in iron and  $\delta$  in the filling material) can still be regarded as belonging to the centre limb, then for the increase in  $L$  as a result of the elastic deformation, we have:

$$\Delta L_{\text{elast}} = \Delta \delta + \frac{L - \delta}{O_1 E_1} F_1 \quad \dots \dots \dots (7)$$

or, approximately, since  $\delta \ll L$ :

$$\begin{aligned} \Delta L_{\text{elast}} &= \Delta \delta \left( 1 - \frac{L}{O_1 E_1} S_1 \right) = \\ &= -\frac{B^2}{2\mu_0} \left( \frac{O_1}{S_1 + S_2} \right) \left( 1 - \frac{L}{O_1 E_1} S_1 \right) \quad \dots \dots \dots (8) \end{aligned}$$

Superimposed upon the elastic deformation is the magnetostrictive deformation, which according to equation (1) is

$$\Delta L_{\text{magn}} = \epsilon L = \frac{\alpha}{2\mu_0} B^2 L \quad \dots \dots \dots (9)$$

for points  $P$  and  $P'$ .

The total increase in the distance  $PP'$  is hence

$$\Delta L = \Delta L_{\text{elast}} + \Delta L_{\text{mag}},$$

from which formula (2) is derived.

*Practical application of equation 2 to hum abatement*

To be able to calculate the desired Young's modulus  $E_2$  of the filling material for the gap, in addition to the given values of  $L$ ,  $\delta$ ,  $O_1$  and  $E_1$ , we must also know the values of  $\alpha$  and  $S_1$ . These latter values must be found by measurement. Firstly, the amplitude of the vibration in the direction  $PP'$  of a coil core of the type shown in fig. 4 but without an air-gap is measured for a known magnetic induction, equal to the value which occurs in practice. In this case from (2) we have:

$$(\Delta L)_{\delta=0} = \frac{B^2}{2\mu_0} L\alpha \dots \dots (10)$$

from which the value of  $\alpha$  for the silicon iron used can be derived. The amplitude is then measured for a coil as in fig. 4 with air-gap but without filling material. Here,  $E_2 = 0$ , and consequently (2) becomes:

$$(\Delta L)_{E_2=0} = \frac{B^2}{2\mu_0} \left[ \alpha L - \frac{O_1}{S_1} + \frac{L}{E_1} \right] \dots (11)$$

The value of  $S_1$  calculated from (11) and of  $\alpha$  from (10) are substituted in (2); then, putting  $\Delta L = 0$ , we have:

$$S_2 = E_2 \frac{O_2}{\delta} = \frac{O_1}{L\alpha} \left( 1 - \frac{E_1 O_1 / L}{S_1} \right) - S_1, \dots (12)$$

and  $E_2$  can now be calculated. A material of roughly this value can be chosen as filler, an additional correction being available in the choice of the cross-section  $O_2 (\neq O_1)$  of the filling.

The arrangement that was used for making the measurements is shown schematically in fig. 5. The ballast under investigation is suspended on rubber bands, so that it can move freely. It is located between the probes of two vibration pick-ups of the type PR 9261 (formerly GM 5526). The vibration pick-ups, which are of the electrodynamic type, are enclosed in boxes of high-permeability alloy, to screen them from the leakage field from the ballast. The arrangement with two pick-ups represents a simple way of eliminating the effects of vibrations of the building. Thanks to the symmetry of the ballast under investigation, if the vibration pick-ups are in contact with symmetrically disposed points on the iron core, they will give voltages of the same phase and amplitude, since such points vibrate in phase and with the same amplitude (in opposite directions). Disturbances from outside, on the other hand, have an equal but opposite effect on the two vibration pick-ups, and may therefore be effectively eliminated by averaging the voltages delivered by the two pick-ups ( $M$  in fig. 5).

The electrodynamic vibration pick-up functions in such a way that the mean voltage is a measure of the velocity amplitude at the points investigated. What we wish to know, however, is the amplitude of the displacement, and it is therefore necessary to integrate the voltage obtained. This is done with the aid of an apparatus of the type PR 9250 (formerly GM 5522). The voltage from the latter ( $I$  in fig. 5) is applied to a cathode-ray oscilloscope via a selective amplifier (which passes only a narrow frequency band at 100 c/s) and via an electronic switch. The current in the coil of the ballast investigated is likewise supplied to the oscilloscope via the second channel of the electronic switch, this making it possible to see whether an expansion or a contraction occurs at the point investigated, at moments when the magnetic induction is at a maximum. To calibrate the apparatus, the amplitude on the oscilloscope screen is compared with a known alternating voltage of 100 c/s (not of the usual frequency of 50 c/s, because of the selective amplifier.)

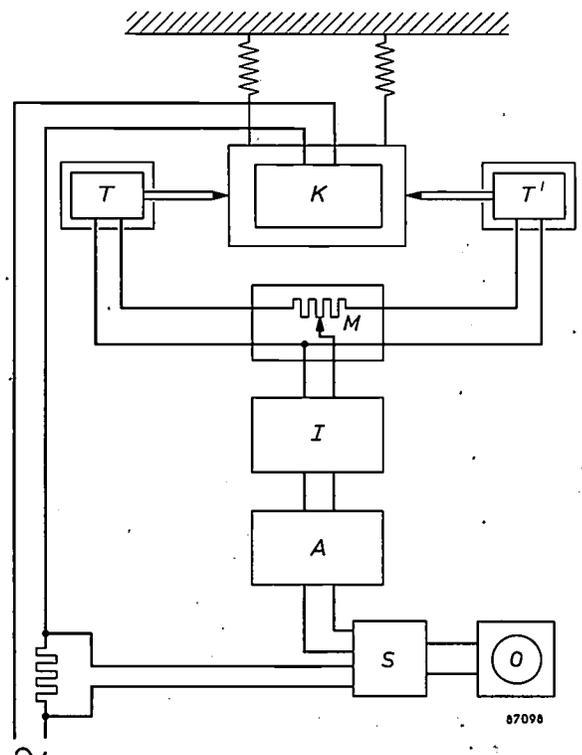


Fig. 5. Apparatus for measuring the vibration amplitude of an elastically suspended ballast  $K$ .  $T$  and  $T'$  electrodynamic vibration pick-ups, whose moving parts have a mass of about 10 grams. The pick-ups exert a constant pressure on the iron core of the ballast of about 8 N (1 newton is roughly equivalent to the weight of 100 grams). The greatest acceleration that can be measured, therefore, amounts to about 800 m/sec<sup>2</sup>. The output voltages from the two pick-ups, added via the potentiometer  $M$ , are integrated (integrator  $I$ ), amplified (in  $A$ ) and displayed alternately with the current from the ballast on a cathode-ray oscilloscope  $O$ , via an electronic switch  $S$ .

With the aid of the arrangement described, the vibration amplitudes (of the order of 0.1 micron or  $10^{-7}$  m) occurring with symmetrical iron cores can be measured with an accuracy of  $3 \times 10^{-9}$  m. In the case of asymmetrical cores, where disturbances from the outside cannot be eliminated in the manner described, the accuracy of measurement is determined purely by the background disturbances. Only by working during quiet periods (at night for example) can an accuracy be obtained approximating to that which can be attained with symmetrical cores.

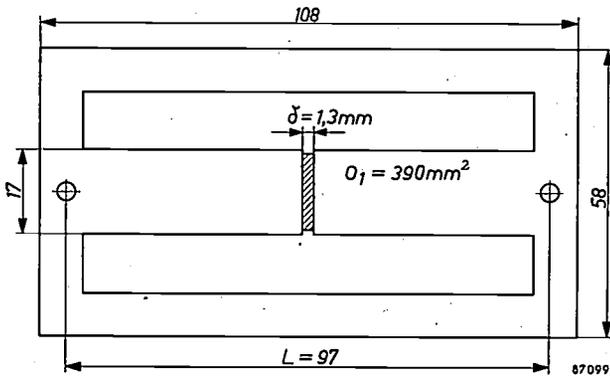


Fig. 6. Dimensioned diagram of the symmetrical core dealt with in the numerical example. The cross-section of the centre limb was  $O_1 = 390 \text{ mm}^2$ .

To conclude this section we may take a numerical example relating to a choke with the dimensions shown in *fig. 6* and with  $E_1 = 2 \times 10^{11} \text{ N/m}^2$ , for use with 40 W lamps, 225 V, 50 c/s. Since the points of attachment  $P$  and  $P'$  are very close to the outside edge, the required amplitudes  $\Delta L$  are virtually identical to the amplitudes measured at this outside edge.

Measurements on a core without air-gap gave  $(\Delta L)_{\text{mag}} = 0.108$  microns for  $B = 1.25 \text{ Wb/m}^2$ , so that  $\alpha = 1.8 \times 10^{-12} \text{ m}^2/\text{N}$ . Measurements on a core with an air-gap, but without filling in the gap gave  $(\Delta L)_{E_1=0} = -0.096$  microns for  $B = 1.25 \text{ Wb/m}^2$ , from which it follows that  $S_1 = 4.9 \times 10^{10} \text{ N/m}^2$ . Substitution of the values found for  $\alpha$  and  $S_1$  in (12), gives  $E_2 O_2 / O_1 = 1.4 \times 10^9 \text{ N/m}^2$ . Such a low value for  $E_2$  cannot, of course, be obtained with metals ( $E \approx 10^{11}$ ), but can be obtained with materials like pressed insulating paper.

*Asymmetric core with air-gap*

In mass-production, choke coils of symmetrical design are sometimes less desirable for reasons of economy. A design frequently encountered in practice, is that shown in *fig. 7a*. Here we must take into account movements of  $P$  and  $P'$  (again the points of attachment) vertical to  $PP'$  as well as parallel to it.

Movements along  $PP'$  can be treated in exactly the same way as with the symmetrical construction. Thus, here also the ideal value can be found for Young's modulus of the gap-filling material. Clearly,  $S_2$  will be appreciably smaller in this case than with the construction shown in *fig. 4*. The ideal filling does not, however, eliminate movements of  $P$  and  $P'$  at right angles to the line  $PP'$ ; for the distance  $PP'$  remains the same, but the limb  $AA'$  expands owing to the preponderant magnetostriction. This means that bending will occur as indicated by the dotted line (which represents the centre line) in *fig. 7a* <sup>8)</sup>.

It can be seen that the distance from  $PP'$  to the centre of gravity of the whole (which lies somewhere on  $MM'$ ) can vary. Moreover, the centre of gravity is not rigidly fixed, since the coil is not able to execute exactly all the movements made by the parts it encloses. This effect certainly cannot be ignored: the mass of the coil forms roughly half the total mass and must thus be taken into account; also the attachment of the coil is not completely free from play. On the average, the play is greater than the vibration amplitude of the core, so that it is a matter of chance where coil and core touch. The

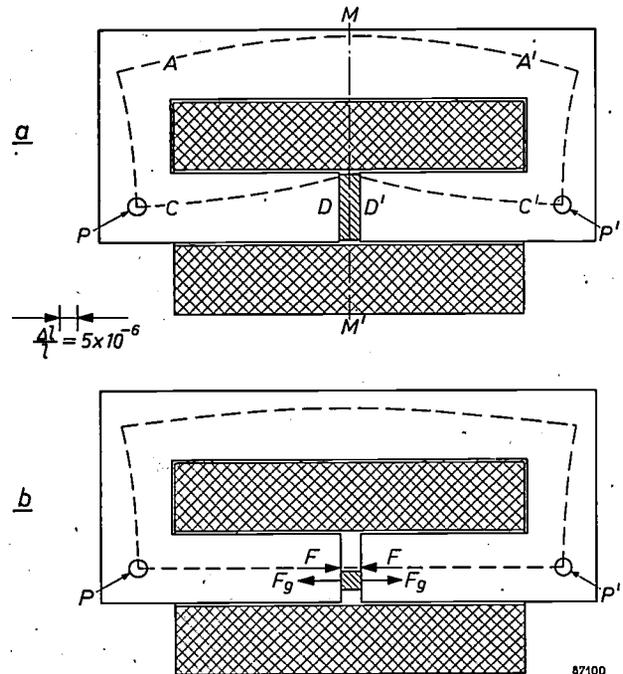


Fig. 7. Deformation pattern of a choke with an asymmetrical core. a) With the air-gap  $DD'$  uniformly filled. The distance  $PP'$  remains constant, but owing to the bending of the limbs  $AA'$  and  $CC'$  the centre of gravity is displaced along the line  $MM'$ . b) With the air-gap  $DD'$  asymmetrically filled. The bending of limb  $CC'$  is eliminated.

<sup>8)</sup> The shape of the dotted line was derived from measurements on a ballast, whose coil consisted of two parts, one on each half of the limb bearing the coil. The iron core was thus accessible at the air-gap.

mutual movement of coil and core differs, therefore, from case to case, and has the unfortunate consequence, that the amplitude of the vibration of  $P$  and  $P'$  perpendicular to the line  $PP'$  has no constant, calculable value. It is only possible to give a maximum value.

The shape of the dotted line in fig. 7a suggest that one means of countering the mutual movement of core and coil might be by applying a bending torque, proportional to the force of magnetic attraction between the poles  $D, D'$ , acting so as to exactly eliminate the bending of the halves of the limbs, described above (there is no objection to limb  $AA'$  bending). This has been found to be possible by filling the air-gap, not uniformly with the material with the exact Young's modulus, required by the condition that  $\Delta L = 0$ , but by filling it with a material with a greater Young's modulus on the outside and a material with a smaller Young's modulus on the inside. The filling can be adjusted so that the condition  $\Delta L = 0$  is still satisfied, but since the point of application of the resulting force  $F_g$  is displaced towards the outside, this force, together with the force of magnetic attraction  $F$  of the poles (acting in the centre line) forms the desired bending torque. The optimum arrangement is best found by experiment. With a choke in which  $\Delta L = 0$  and in which the artifice just described has been applied, the coil will not be able to move at right angles to  $PP'$ , irrespective of where the points of contact are between coil and core (fig. 7b).

Naturally, if  $P$  and  $P'$  are chosen as points of attachment, the movement of the centre of gravity perpendicular to  $PP'$  will not be completely eliminated, for  $AA'$  will still be bent. It might seem worth while to see whether any improvement can be gained from a slight over-compensation of the movement of the coil with respect to the core; experience shows, however, that even with the best attachment of the coil to the core obtainable in mass production, such a process again leads to excessive differences in the resulting vertical component of the individual coils. Fortunately, however, the reduction in hum attained with the compensation described is already very satisfactory.

#### *Cores without air-gaps (transformers)*

Where the mains voltage is low (110 or 127 V) the ballasts usually consist of leakage transformers, i.e. of a transformer in combination with a choke. Sometimes, however, small transformers are also used separately. This is the case, for example, with type "TL" S, where normally an incandescent lamp functions as ballast<sup>9)</sup>, but where, in addition, a transformer is necessary if a mains voltage of 220 V is not available.

Since these transformers do not possess air-gaps, the process outlined above for combating hum cannot be applied.

In principle two solutions are feasible:

<sup>9)</sup> W. Elenbaas and T. Holmes, Philips tech. Rev. 12, 129-135, 1950/51.

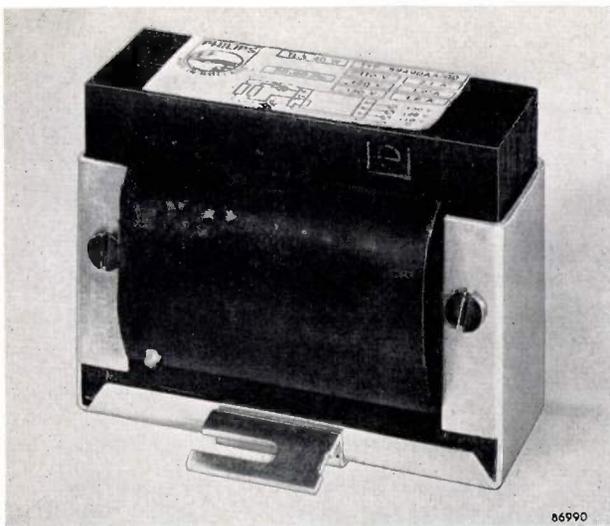


Fig. 8

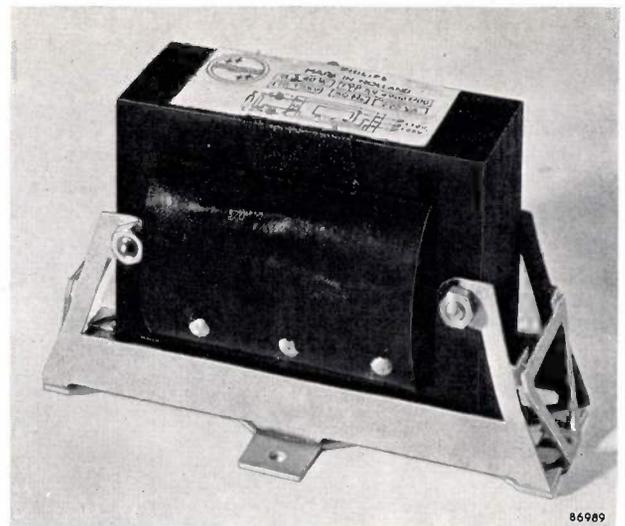


Fig. 9

Fig. 8. Transformer in a rectangular bracket with flexible corners. The bracket is screwed to the mounting surface through the middle of the horizontal connecting piece.

Fig. 9. Transformer in a rigid double bracket with convergent arms.

- a) The use of transformer iron which shows hardly any magnetostriction.
- b) Attachment of the transformer exactly at its centre of gravity.

In practice neither solution is feasible. "Free" suspension (on springs) is also out of the question.

It has however been found that if only two points suitable for attachment can be found on the iron core, which vibrate along the same line in phase and with equal and opposite amplitudes, a bracket can be constructed, which while being sufficiently firm, does not transmit vibrations to the ceiling or wall.

Two types of bracket have been developed, both for symmetrical transformers; these are shown in figs. 8 and 9.

The type shown in fig. 8 is very simple. The body and arms of the right-angled bracket are reinforced with flanges; the corners are purposely left flexible. The core of the transformer may be assumed here to be the same as that of the choke in fig. 4, but without an air-gap. Here, apart from the symmetry of the vibrations mentioned earlier, use is also made of the shape of the deformed iron core.

The shape of the core, as measured in the case of the transformer for which this bracket was designed (for  $B_{max} = 1.25 \text{ Wb/m}^2$ ), is given in fig. 10. The central parts of the yokes of the core, between  $X$  and  $Y$  or  $X'$  and  $Y'$ , bend but little and move almost

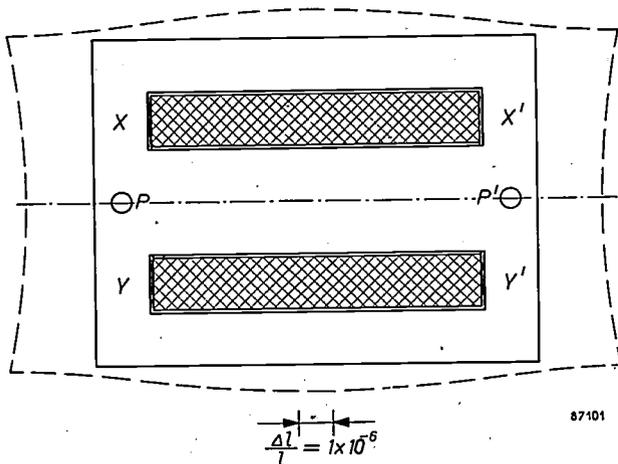


Fig. 10. Deformation pattern of a transformer with symmetrical core, of the type shown in figs. 8 and 9. The dotted line indicates the outline of the deformed core.

parallel to each other. If the arms of the bracket are clamped sufficiently firmly to the core at these places (and this is possible even with a relatively low coefficient of friction), they will then be displaced parallel to each other. This means that the connecting piece of the bracket will be stretched without bending. The centre of the connecting piece thus remains

stationary and can function as the point of attachment of the whole transformer to a wall, etc.

The second type, a convergent double-bracket, is represented diagrammatically in fig. 11, which shows only the elastically neutral line of the body

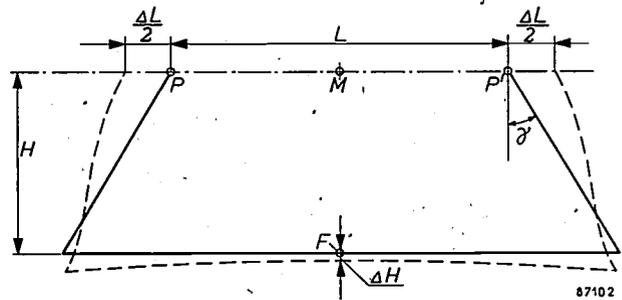


Fig. 11. Diagram of the deformation of the halves of the bracket shown in fig. 9.

and the arms (for one half of the bracket). The way in which the ends of the arms of the two halves of the bracket are attached to the transformer core can be seen in the photograph, fig. 9. The centre  $F$  of the connecting piece of each half of the bracket is screwed on to the wall to which the transformer is to be affixed. In a bracket of the type shown, which has rigid corners, substantial bending of the connecting piece will occur, and in general there will be a displacement  $\Delta H$  of  $F$  with respect to the centre of gravity  $M$  of the transformer (which, on account of the relatively low mass of the bracket, practically coincides with the centre of gravity of the whole system). However, if the bracket has suitable dimensions  $\Delta H = 0$ , so that no vibration will be transmitted to the wall, i.e. no hum will be audible.

At moments when the induction  $B$  is a maximum, points  $P$  and  $P'$  of the core in fig. 11 are both displaced towards the outside through a distance  $\frac{1}{2}\Delta L$ . Provided the transformer is sufficiently firmly attached to the bracket, the extremities of the arms move with  $P$  and  $P'$ . The shape then assumed by the neutral line, is indicated (exaggerated) by the dotted line in fig. 11. Calculation of the elastic deformation gives the following expression for the displacement  $\Delta H$  from the centre  $F$ :

$$\frac{\Delta H}{\Delta L} = \frac{\frac{3}{4} \left(\frac{L}{H}\right)^2 \cos \gamma + \left(\frac{L}{H}\right) \sin \gamma - \sin \gamma \operatorname{tg} \gamma - \frac{I_1}{I_2} \operatorname{tg} \gamma}{4 \left(\frac{L}{H}\right) \cos \gamma + 4 \sin \gamma + \frac{I_1}{I_2}} \quad (13)$$

$I_1$  and  $I_2$  are the linear moments of inertia or second moments of area<sup>10)</sup> of the cross-sections of the body and the two arms respectively; for the significance of  $L$ ,  $H$ ,  $\gamma$  see fig. 11.

<sup>10)</sup> For the relevance of the second moment of area  $I = \int y^2 dA$  of a cross-section to the elastic strain, the reader is referred to text-books on elasticity. For a rod of rectangular cross-section of breadth  $b$  and length  $h$ , the axis running centrally across the area, parallel to  $b$ ,  $I = bh^3/12$ .  $I$  is expressed in  $\text{inch}^4$  or  $\text{m}^4$  (here, the latter).

It follows from formula (13) that  $\Delta H$  can be either positive or negative and can also be made equal to zero. For the latter, the condition:

$$\frac{I_1}{I_2} = \frac{1}{4} \left( \frac{L}{H} \right)^2 \frac{\cos^2 \gamma}{\sin \gamma} + \left( \frac{L}{H} \right) \cos \gamma - \sin \gamma \quad (14)$$

must be satisfied. For given values of  $L$ ,  $H$  and  $\gamma$  the desired value of  $I_1/I_2$  can be read off from fig. 12, in which the ratio  $I_1/I_2$  is plotted against  $L/H$  (both scales logarithmic), for various values of  $\gamma$ .

It should be remarked that equation (14) is valid only when the natural frequencies of the equipment are high compared with the vibration frequency of the transformer (100 c/s). This condition can be easily satisfied, since a condition is imposed only on the ratio  $I_1/I_2$ , while  $I_1$  or  $I_2$  may still be chosen at will.

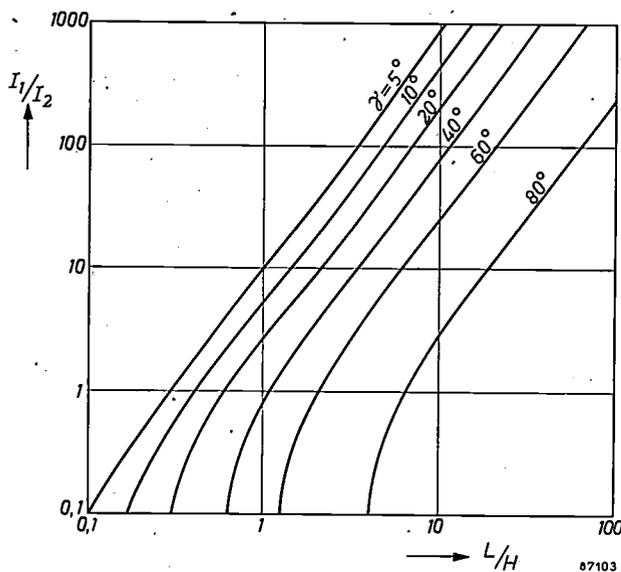


Fig. 12. Graph giving the desired value of  $I_1/I_2$  as a function of  $L/H$ , for various values of the angle  $\gamma$  (see fig. 11).

### Measures to reduce rustle

We have seen that hum is the consequence of vibrations set up in the core of the ballast, which are transmitted to the wall.

The core can also vibrate freely (to be compared with a freely vibrating tuning fork). Owing to the large cross-section, these vibrations are of high frequency (several kc/s). Moreover, since the core is built up of laminations, they are strongly damped. The vibrations are produced in an irregular manner by the small current peaks described earlier, which are superimposed on the approximately sinusoidal current. Owing to the indeterminacy of the phases and the spread in the frequency of these vibrations, they give rise to a noise spectrum known as rustle.

It is evident that rustle will occur only with choke coils. With transformers the magnetization is substantially sinusoidal owing to the low impedance of the mains shunting the primary winding.

At the frequencies involved with rustle, the transmission of vibrations to the walls is found to be very slight<sup>11)</sup>. This is fortunate, since the mounting methods for countering hum discussed above, can generally offer no solution: the calculations made in connection with these methods were based on the assumption that the forces and the deformations they produce are in phase. This holds well enough at low frequencies, but not at the high frequencies associated with rustle, which lie within the range of the natural frequencies of the equipment.

The chief problem, therefore, is to prevent direct emission of sound. This can be best achieved by surrounding the ballast by a closed box; however, other problems then arise. In connection with the measures against hum, the equipment should be mounted in the box only at vibration-free points (nodal points), yet to obtain sufficient dissipation of the heat developed the box should be packed with some heat-conducting mass. The filling necessarily forms a mechanical contact between the walls of the box and the vibrating parts of the equipment, and must therefore be not only a good conductor of heat but also a good insulator against vibrations. For vibrations in the frequency range of rustle it would be sufficient that the equipment nowhere makes direct contact with the box. For hum frequencies, which are more difficult to insulate, the fact that on opposite sides of the nodal points the displacement is in opposite directions has the result that the net transmission of vibrations by the filling to the box and thence to the mounting surface is sufficiently attenuated.

A common solution is to float the ballast, as it were, in a bituminous mass. To improve heat conduction, sand is frequently added to the bitumen. A difficulty here is that whilst the bitumen filler should preferably be soft and pourable at a temperature high above the working temperature of the coils, the electrical insulation cannot withstand such a temperature during manufacture. For this reason, a filling material has to be chosen which softens at relatively low temperatures, with the attendant risk that sooner or later the ballast will sink and come into contact with the metal wall of the box, greatly to the detriment of the acoustical insulation.

A good solution to the problem has been found by filling the space in the box with corrugated aluminium foil (silver paper). If a suitable "filling factor" be chosen, the heat conductivity is quite as good as or better than that of the bitumen-sand mixture.

<sup>11)</sup> W. Elling Bestimmung mechanischer Eingangsimpedanzen, *Acustica* 4, 396-402, 1954.

Vibrations in the frequency range of rustle are adequately damped, since they must pass many contact points between the sheets of foil, before reaching the wall of the box.

In conclusion, we shall briefly note the behaviour of the filling with respect to the hum vibrations of the iron core. Initially, of course, points on all parts of the iron core will come into contact with the aluminium. Some of these points vibrate with relatively large amplitudes, others with small amplitudes. Where the hum amplitudes are large, the elastic limit of the sheets of foil will be exceeded, so that these sheets will be plastically deformed by the vibrating iron core. In this way room will be made for vibration, the final result being, that the aluminium filling will only be in contact with the iron core where the amplitude of the hum is small or non-existent and the remainder of the core will be left free. This is just what we want, for now the low frequency vibrations are not transmitted via the box to the wall or ceiling.

Proof of this explanation is furnished by a simple

experiment. If the whole apparatus (i.e. complete with box) is put in another position, then owing to the change in direction of the gravitational force, the sheets of foil will come into contact with quite different points on the iron core. A certain increase in hum can indeed often be heard, but the latter rapidly falls once more to its former low and acceptable level.

---

**Summary.** The origin of troublesome noises produced by fluorescent lighting installations lies in the ballasts. The latter vibrate with twice the mains frequency (producing hum) and also execute free vibrations of high frequency (rustle). Both forms of vibration result from the alternating elastic and magnetostrictive deformations due to the variations of current. High-frequency vibrations are directly transmitted to the air, low-frequency vibrations via the wall or ceiling. In ballasts with air-gaps (chokes), hum can be countered by filling the gap with a suitable material (mutual compensation of magnetostrictive and elastic deformation at the points of attachment). In ballasts without air-gaps (transformers) transmission of vibrations to the wall can be avoided by use of suitably designed brackets. Rustle can be combated by enclosing the ballast, surrounded by a sound-insulating material, in a box. The insulating material must however be heat conducting; aluminium foil has been found to answer the purpose.

## THE GYRATOR, AN ELECTRIC NETWORK ELEMENT

by B. D. H. TELLEGEN.

621.372.2

*There is a traditional antithesis between the practical mind and the mind inclined to pure scientific pursuits: the former is mainly interested in knowledge that can be put immediate to use, while the latter ponders fundamentals and tries to introduce greater clarity and generality by rigorous considerations which the practical man probably finds superfluous. The invention of the gyrator by Prof. Tellegen is a striking illustration of the service that can be rendered to engineering by the pure scientific approach: originally regarded as a hypothetical possibility, a possibility that had to be recognized for the sake of completeness, the gyrator has subsequently and perhaps rather surprisingly become a reality in the world of microwaves.*

Wide use is made in electrical engineering of networks composed of resistors, coils and capacitors, the latter being known as network elements. In use, there is conversion and exchange of energy within and between these network elements. In resistors electrical energy can be converted into heat; in coils and capacitors energy can be stored and later released. The networks are provided with terminal pairs allowing an exchange of energy to take place with the exterior. A terminal pair consists of two conductors between which a voltage of instantaneous value  $v$  can exist and through which a current of instantaneous value  $i$  can flow (see *fig. 1*). The energy

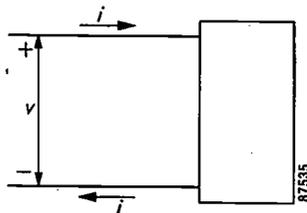


Fig. 1. Network with one terminal pair.

supplied to the network via the terminal pair in a time  $dt$  is  $ivdt$ ; it can be positive or negative. The network determines relations between the voltages and currents at the terminals, the number of relations being equal to the number of terminal pairs. If for example the network consists of one resistor  $R$ , it has one terminal pair and the relation is  $v = Ri$ .

### Networks made up of resistors, coils and capacitors

For the user networks are characterized primarily by the relations that hold between the terminal voltages and currents. As long as the relations between these voltages and currents are the ones desired, the actual composition of the network is, for the user, of secondary concern. It is therefore of great importance to know what sets of such relations

are possible with networks consisting of resistors, coils and capacitors. All the possible sets of relations form an arsenal from which the user can take his choice.

In order to build up the whole arsenal, i.e. a complete list of the possible sets of relations, let us first look for general properties of the relations; we can then try to demonstrate that, for each set of relations having these properties, it is possible to design a network of resistors, coils and capacitors for which the said relations are valid. This can be put in another way by saying that we shall try to find necessary and sufficient conditions for the sets of relations if they are to be realizable by networks composed of resistors, coils and capacitors.

General properties of the relations are that they consist of *linear* differential equations relating the terminal voltages and currents, and that the coefficients of the equations are *constant*, i.e. not dependent on time but determined only by the magnitude of the elements composing the networks. Furthermore, the networks contain no source of energy; they are said to be *passive*. Certain properties of the coefficients of the differential equations can be deduced from the fact that the networks are passive. Let us consider, for example, a network with one terminal pair, the properties of which are governed by a differential equation of the first order, so that we can write:

$$a \frac{di}{dt} + bi = c \frac{dv}{dt} + dv, \quad \dots \quad (1)$$

where  $a$ ,  $b$ ,  $c$  and  $d$  are constants. From the passivity of the network it can be shown that  $a$ ,  $b$ ,  $c$  and  $d$  all have the same sign.

This can be demonstrated as follows. If we short-circuit the network, in other words if we keep  $v = 0$ , the current will be determined by:

$$a \frac{di}{dt} + bi = 0.$$

The solution of this differential equation is:

$$i = C e^{-\frac{b}{a} t}$$

where  $C$  is the constant of integration. In consequence of the passivity, the current in the short-circuited network cannot rise indefinitely and hence  $a$  and  $b$  have the same sign. It follows in a similar way, by considering the network to be open-circuited, i.e. by keeping  $i = 0$ , that  $c$  and  $d$  have the same sign.

For direct current and voltage equation (1) simplifies to

$$bi = dv,$$

so that  $b/d$  represents the DC resistance of the network. This must be positive on account of the network's passivity, and hence  $b$  and  $d$  also have the same sign.

For convenience we shall group the three properties together and say that the sets of relations characterizing networks made up of resistors, coils and capacitors are linear, constant, passive. As far as networks with one terminal pair are concerned, these properties are not only necessary but also sufficient, for Brune has demonstrated<sup>1)</sup> that any linear constant passive relation can be realized by a network made up of resistors, coils and capacitors. In particular, it can be demonstrated that any relation having the form of equation (1) can be realized by a network consisting either of two resistors and a coil (if  $a/c > b/d$ ), or of two resistors and a capacitor (if  $a/c < b/d$ ). The two networks are shown in fig. 2.

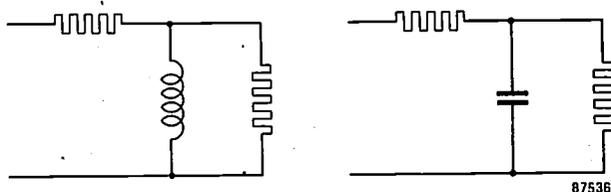


Fig. 2. Networks with one terminal pair and of the first order.

For any network consisting of resistors, coils and capacitors and having more than one terminal pair a set of relations holds that, besides being linear constant passive, has the property known as reciprocity. To explain this property we shall take a network with two terminal pairs and express the two terminal voltages in terms of the two terminal currents; for the present purpose it will be convenient to write the relationships in complex form, rather than use the instantaneous values. Accordingly, we obtain:

$$\begin{cases} V_1 = Z_{11}I_1 + Z_{12}I_2, \\ V_2 = Z_{21}I_1 + Z_{22}I_2, \end{cases} \dots \dots \dots (2)$$

where  $I_1, I_2, V_1$  and  $V_2$  represent the complex values of the terminal currents and voltages. If we assume the sign convention for currents and voltages as indicated in fig. 3, it can be shown that always

$$Z_{21} = Z_{12} \dots \dots \dots (3)$$

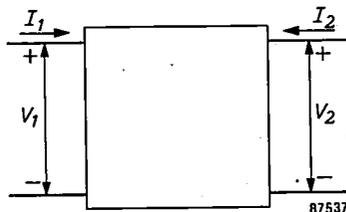


Fig. 3. Network with two terminal pairs.

Expressing the current  $I_1$  through the first terminal pair and the voltage  $V_2$  across the second terminal pair in terms of the two other terminal quantities, we find from equations (2):

$$\begin{cases} I_1 = \frac{1}{Z_{11}} V_1 - \frac{Z_{12}}{Z_{11}} I_2, \\ V_2 = \frac{Z_{21}}{Z_{11}} V_1 + \frac{Z_{11}Z_{22} - Z_{12}Z_{21}}{Z_{11}} I_2. \end{cases} \dots \dots \dots (4)$$

The coefficient of  $V_1$  in the second equation above is now equal and opposite in sign to the coefficient of  $I_2$  in the first equation. The relation (3) between two coefficients of equations (2) and the relation just discussed between two coefficients of equations (4) are known as the reciprocity relations.

For networks with more than one terminal pair Bayard and others have demonstrated<sup>2)</sup> that any linear constant passive set of relations having the reciprocity property can be realized by a network consisting of resistors, coils and capacitors. Hence the said properties of sets of relations are sufficient as well as necessary.

Our arsenal of all possible sets of relations capable of realization by networks made up of resistors, coils and capacitors is thus complete.

Linear constant passive systems

The result obtained above is not entirely satisfactory, however. We have based our considerations on the resistors, coils and capacitors that had their origin in the laboratory. Electrical engineering has simply accepted these elements for constructing networks. There is something arbitrary and fortuitous about all this. Must we necessarily use these

<sup>1)</sup> O. Brune, J. Math. Phys. 10, 191, 1931.

<sup>2)</sup> M. Bayard, Bull. Soc. franç. Elect. 9, 497, 1949. Also B. D. H. Tellegen, J. Math. Phys. 32, 1, 1953, which gives further references to the literature.

elements for the building-up of networks? Should an affirmative answer be found to this question, we would be faced with another one: are other elements conceivable, elements that have not been found in the laboratory?

In order to investigate these questions we must not proceed from the basis of the conventional network elements and the networks composed of them but approach the problem from another direction. We must consider "black boxes", systems with terminal pairs that are characterized only by the sets of relations existing between the terminal voltages and currents without concerning ourselves with what is inside them. In what follows we shall confine ourselves to linear constant passive systems, by which we mean systems characterized by linear constant passive sets of relations. We could of course subject ourselves to less stringent limitations: we could also consider systems containing energy sources (active systems, such as amplifiers), systems characterized by linear equations with variable coefficients (variable systems), or systems characterized by non-linear equations (non-linear systems). However, such systems have properties that depart appreciably from those of the networks considered above, and are of a more complex nature. We shall therefore leave them aside.

Let us now try to find the "simplest" kinds of linear constant passive system. These simplest systems we shall call network elements, by *definition*. In this way we shall try to arrive at a set of network elements such that any linear constant passive system can be realized as a network composed of them. We can regard such a set of network elements as a complete set.

In order to give a meaning to the epithet "simplest", we shall have to classify our "black boxes". We shall do so according to the number of terminal pairs, the order of the differential equations characterizing them, and whether or not they are able to dissipate electrical energy, i.e. to transform it into heat. This leads, as a first step, to the investigation of systems with one terminal pair, of zero order and involving no dissipation. For such systems the power supplied must be zero at any instant, i.e.  $iv = 0$ ; therefore either  $i = 0$  (open terminal pair) or  $v = 0$  (short-circuited terminal pair). This does not produce a network element. In accordance with the threefold classification, our next step will be to investigate three classes of systems, namely (1) systems with one terminal pair, of zero order and involving dissipation, (2) systems with one terminal pair, of the first order and involving no dissipation, and (3) systems with two terminal

pairs, of zero order and involving no dissipation.

- (1) Systems with one terminal pair, of zero order and involving dissipation

These systems are characterized by an equation having the form:

$$v = Ri, \quad R > 0. \quad \dots \quad (5)$$

The fact that  $R$  is positive is a consequence of the passivity of the network, this requiring that  $iv$  should be positive.

- (2) Systems with one terminal pair, of the first order and involving no dissipation

Systems with one terminal pair and of the first order are characterized by an equation of the form of (1). It can be shown that this represents a system involving no dissipation either if  $b = 0$  and  $c = 0$  or if  $a = 0$  and  $d = 0$ . Hence these systems are of two kinds. The first kind is characterized by an equation of the form:

$$v = L \frac{di}{dt}, \quad L > 0, \dots \quad (6)$$

and the second kind by an equation of the form:

$$i = C \frac{dv}{dt}, \quad C > 0. \dots \quad (7)$$

The above may be derived as follows. The system to which (1) is applicable has an impedance of

$$Z = \frac{j\omega a + b}{j\omega c + d}.$$

In order that there should be no dissipation, the real part of  $Z$  must be zero at all frequencies. It follows that  $ac = 0$  and  $db = 0$ . Since  $a = 0$  and  $b = 0$  means that  $Z = 0$ , and  $c = 0$  and  $d = 0$  means that  $Z = \infty$ , these solutions may be disregarded. Thus either  $c$  and  $b$ , or  $a$  and  $d$ , must be zero, and these two cases produce equations (6) and (7), respectively. Since the passivity of the network requires that  $a$ ,  $b$ ,  $c$  and  $d$  should all have the same sign (see above), it further follows that  $L = a/d$  and  $C = c/b$  are both positive.

Before examining the third class of simplest systems we shall take a closer look at the three systems, (5), (6), and (7), already found. As stated above, we shall regard them as network elements. The latter, then, are defined by equations, not by the physical means required to realize them. For example, we cannot infer from the equation  $v = L di/dt$  that it describes a "coil". A superconductor has the same equation when we take the effect of the mass of the conducting electrons into account. The kinetic energy of these electrons then takes the place of the magnetic energy of the coil.

The three network elements, so defined, are "ideal"; they can only be realized approximately. For example, (5) is approached by a resistor with low stray capacitance and low self-inductance (6) by a low-loss coil with a low stray capacitance, and (7) by a low-loss capacitor with a low self-inductance (hence the symbols  $R$ ,  $L$  and  $C$  used in the equations).

The way in which the network elements are realized is only of secondary importance to the user; it is the external properties given by (5), (6) and (7), that are of primary interest. This remark is similar to that made earlier regarding networks: the user has little interest in their internal composition.

(3) Systems with *two terminal pairs*, of zero order and involving no dissipation

For these systems the total power supplied via the terminal pairs is zero at any instant, so that

$$i_1 v_1 + i_2 v_2 = 0,$$

where  $i_1$ ,  $i_2$ ,  $v_1$  and  $v_2$  represent instantaneous values of the terminal currents and voltages, the sign convention for these quantities being assumed to be in accordance with fig. 3. This condition results in two kinds of systems, namely those characterized by equations of the form:

$$\left. \begin{aligned} i_1 &= -ni_2, \\ v_2 &= nv_1, \end{aligned} \right\} \dots \dots \dots (8)$$

and those characterized by equations of the form:

$$\left. \begin{aligned} v_1 &= -si_2, \\ v_2 &= si_1, \end{aligned} \right\} \dots \dots \dots (9)$$

Equations (8) and (9) may be derived as follows. If, for a two-terminal-pair system of zero order, we express the two terminal voltages in terms of the two terminal currents, we can write the equations as:

$$\left. \begin{aligned} v_1 &= a_{11}i_1 + a_{12}i_2, \\ v_2 &= a_{21}i_1 + a_{22}i_2. \end{aligned} \right\} \dots \dots \dots (10)$$

It follows from this that:

$$i_1 v_1 + i_2 v_2 = a_{11}i_1^2 + (a_{12} + a_{21}) i_1 i_2 + a_{22}i_2^2. \dots (11)$$

In order that (10) should represent a system involving no dissipation, (11) must be zero for all values of  $i_1$  and  $i_2$ ; hence  $a_{11} = 0$ ,  $a_{12} + a_{21} = 0$ , and  $a_{22} = 0$ , which results in (9).

Starting from the equations that express  $i_1$  and  $i_2$  in terms of  $v_1$  and  $v_2$  we also arrive at (9). However, it is also possible to conceive systems in which  $v_1$  and  $v_2$  cannot be expressed in terms of  $i_1$  and  $i_2$ , or vice versa. This is the case for systems that are characterized by a relation between  $i_1$  and  $i_2$  and a relation between  $v_1$  and  $v_2$ , making it impossible to choose  $i_1$  and  $i_2$  or  $v_1$  and  $v_2$  as independent variables. In order to investigate these systems we can start from equations expressing  $i_1$  and  $v_2$  in terms of  $v_1$  and  $i_2$ , or vice versa. We then arrive at (8).

We shall regard the systems thus found, (8) and (9), as two further network elements. System (8) is called an ideal *transformer* and we have proposed<sup>3)</sup> the name ideal *gyrator* for system (9). Observations similar to those made above on (5), (6) and (7) may also be made on these two systems.

An approximation to the ideal transformer is given by two tightly coupled low-loss coils of large self-inductances. This is in fact the way in which the concept first arose. The theoretical route by which we have now arrived at the same concept shows why we should regard the ideal transformer as a separate network element, defined, in fact, by (8). From the remarks on reciprocity in connection with equations (4) we see that the ideal transformer also possesses this property. The ideal gyrator does *not* possess the property of reciprocity, as will be clear from a comparison of (9) with (2) and (3). It is therefore impossible to obtain an approximation to it by combining resistors, coils and capacitors, because such a combination always has the property of reciprocity, as we saw above. For the realization of the gyrator other physical means are necessary. One way of realizing it, making use of gyromagnetic effects in ferromagnetic materials, will be described in a further article in this Review<sup>4)</sup>.

The addition of the gyrator to the set of network elements has extended the list of possible sets of relations between terminal voltages and currents. This extension was made possible by the fact that we started from linear constant passive systems as such, without imposing on them the condition of reciprocity. The fact that system (9), while being linear constant passive lacks the property of reciprocity, shows that the latter is not a consequence of linearity, constancy and passivity, and that imposing it does indeed constitute a limitation of the possibilities.

Some properties of the gyrator

The ideal gyrator has the property of "gyrating" a current into a voltage, and vice versa. The coefficient  $s$ , which has the dimension of a resistance, we call the gyration resistance;  $1/s$  we call the gyration conductance. We shall represent the gyrator in circuit diagrams by the symbol shown in fig. 4.

The following properties of the ideal gyrator can be easily derived from (9).

<sup>3)</sup> B. D. H. Tellegen, Philips Res. Rep. 3, 81, 1948.  
<sup>4)</sup> H. G. Beljers, Application of ferroxcube to uni-directional waveguides, to appear shortly in this Review.

If we have the output terminals open-circuited, i.e.  $i_2 = 0$ , the input terminals are short-circuited,

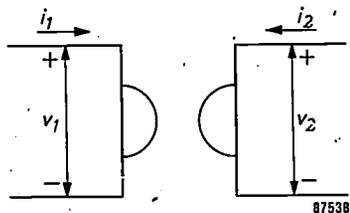


Fig. 4. Symbol for the gyrator.

i.e.  $v_2 = 0$ , and vice versa. If we connect an inductance  $L$  or a capacitance  $C$  across the output terminals, we find a capacitance  $L/s^2$  or, in the second case, an inductance  $s^2C$  between the input terminals. In general, if we connect an impedance  $Z$  across the output terminals, we find an impedance  $s^2/Z$  between the input terminals. An impedance  $Z$  in series or in parallel with the output terminals has the same effect as an impedance  $s^2/Z$  in parallel or in series, respectively, with the input terminals (fig. 5).

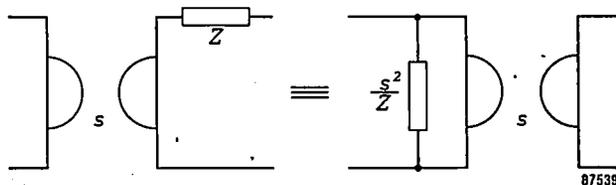


Fig. 5. An impedance in series with one terminal pair of an ideal gyrator is equivalent to another impedance in parallel with the other terminal pair.

Two ideal gyrators in cascade form an ideal transformer; an ideal gyrator and an ideal transformer in cascade form another ideal gyrator.

The combination of a gyrator with a resistor shown in fig. 6a gives a system to which the following apply:

$$\left. \begin{aligned} v_1 &= Ri_1 + (R-s)i_2, \\ v_2 &= (R+s)i_1 + Ri_2. \end{aligned} \right\} \dots \dots (12)$$

Thus an input  $i_1$  gives rise to a voltage component  $(R+s)i_1$  across the output; an output current  $i_2$  gives rise to a voltage component  $(R-s)i_2$  across the input. If  $R = s$ , the latter component is zero, so that the input voltage is independent of the

output current and dependent only on the input current. In such a case we may say that the system transmits only in the direction from input to output, and not in the reverse direction. The system shown in fig. 6b has properties of a similar kind; this can be easily demonstrated by writing down the equations expressing the terminal currents in terms of the terminal voltages.

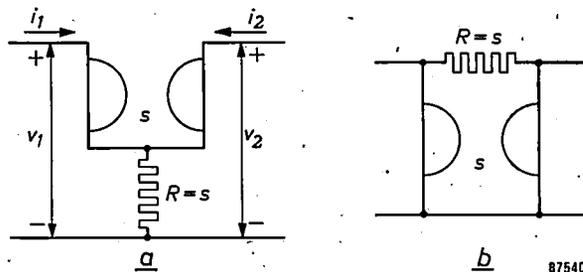


Fig. 6. Uni-directional networks, i.e. networks that transmit only in one direction.

No further elements remain to be added to the five ideal linear constant passive network elements: Oono and Yasuura have demonstrated <sup>5)</sup> that any linear constant passive system can be realized as a network composed of these five network elements. The five ideal network elements therefore form a complete set.

<sup>5)</sup> Y. Oono and K. Yasuura, Mem. Fac. Engng., Kyushu Univ. 14, 124, 1954; also in Ann. Télécomm. 9, pp. 73 and 109, 1954.

**Summary.** The writer considers the electric networks having one or more terminal pairs that can be built up from conventional network elements, namely resistors, coils and capacitors. The set of relations between terminal voltages and currents determined by such a network has properties that correspond to the linear, constant and passive nature of such networks and, furthermore, has the property of reciprocity. It has been demonstrated that, conversely, these properties are sufficient for any set of relation possessing them to be realizable by a network. The writer reverses this line of reasoning and raises the question: what (ideal) network elements must be introduced in order to make it possible to realize all linear constant passive systems. There are no grounds for including reciprocity amongst the properties imposed. It is then shown that, apart from the conventional elements — resistor, coil, capacitor, and the ideal transformer — a new network element has to be introduced; this new element has been christened ideal gyrator. Networks containing the new element generally lack the property of reciprocity. A brief sketch is given of some of the properties of the gyrator.

## LEVITATION BY STATIC MAGNETIC FIELDS

531.51:538.12

Levitation of objects is a problem that has fascinated philosophers throughout the ages. A solution of this problem may have practical importance, e.g. for eliminating friction in rotation or horizontal displacement, or for avoiding chemical reactions between the object and its surroundings. With an eye to the practical importance of levitation we feel justified here in disregarding those aspects of it associated with magic, spiritualism and psychic phenomena, and confining ourselves to a few remarks concerning its physical realization and, more particularly, to a description of some informative experiments in this field.

The problem is that of compensating the gravitational force exerted on a body in such a way that, *in the absence of any mechanical contact*, it is held in stable equilibrium with respect to the earth. Probably the most spectacular solution is that where the compensating forces are generated by the body itself as reactive forces, in accordance with the rocket principle. A feature of the rocket solution is that it is in principle possible to levitate a body at *any* height above the earth. For every pound of weight lifted, however, an enormous power is required, so much indeed that using conventional fuels levitation cannot be maintained for more than a few minutes. The use of nuclear power may be calculated to raise the theoretically attainable maximum duration to about 250 days <sup>1)</sup>.

When the gravitational forces exerted on the body are compensated by external forces, such as those exerted by electric or magnetic fields, levitation is possible only at relatively short distances from the equipment that generates these forces and which are themselves earthbound. In these cases, though, considerably less power is required and levitation can be maintained for an unlimited period. In this connection we would mention a spectacular experiment carried out in 1939, in which an aluminium disc of about a foot in diameter was made to float freely in an alternating magnetic field, generated by a special magnet fed with alternating current from the mains <sup>2)</sup>. Here the gravitational force was compensated by the force exerted by the magnetic field

on the electrons moving in the conductor. An analogous arrangement was employed some years ago for melting metals in vacuo without a crucible <sup>3)</sup> — a technique whose potentialities might prove to be important.

The essential feature of the above experiments is that the levitated body is kept in *stable* equilibrium.

According to a well-known theorem, due to Earnshaw, this cannot be the case with a magnetic body floating freely (i.e. without any mechanical support) in a *constant* magnetic field <sup>4)</sup>. It is evident, for example, that a small permanent magnet  $M_2$  suspended below a large permanent magnet  $M_1$  such that the gravitational force on  $M_2$  is exactly compensated (*fig. 1*) cannot be in stable equilibrium. The slightest upward displacement of  $M_2$  results in a greater magnetic attraction and  $M_2$  continues

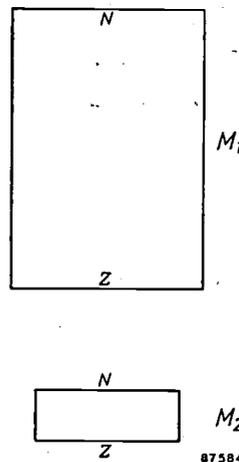


Fig. 1

to rise until it clings to  $M_1$ ; any slight downward displacement of  $M_2$  reduces the magnetic attraction and  $M_2$  then continues to fall. This unstable equilibrium can, however, be turned into a stable one if the permanent magnet  $M_1$  is replaced by an electromagnet, the excitation current of which is automatically regulated in accordance with the position of  $M_2$ . An elegant set-up for this purpose is to arrange that the position of  $M_2$  controls the value of a capacitance, which in its turn controls the excitation of the magnet. In an experimental arrangement, a body weighing 210 kg (about 4 cwt) was levitated in this way <sup>5)</sup>.

The power required in the above case was remarkably low: only 1.3 watts per kg. However, levitation may be effected by a static magnetic field, consuming no energy at all. Earnshaw's theorem, according to which a magnetic body floating in a static magnetic field cannot be in a state of stable equilibrium, is invalid for substances with a relative

<sup>1)</sup> For this calculation, and more generally for a survey of the theoretical and practical possibilities of levitation, reference should be made to: A. H. Boerdijk, Technical aspects of levitation, Philips Res. Rep. 11, 45-56, 1956.

<sup>2)</sup> B. D. Bedford, L. H. B. Peer and L. Tonks, Gen. El. Rev. 42, 246-247, 1939.

<sup>3)</sup> E. C. Okress, D. M. Wroughton, G. Comenetz, P. H. Brace and J. C. R. Kelly, J. appl. Phys. 23, 545-552, 1952.

<sup>4)</sup> S. Earnshaw, Trans. Cambr. Phil. Soc. 7, 97-112, 1842. See also: J. C. Maxwell, A treatise on electricity and magnetism, Clarendon, Oxford 1873, part I, pp 139-141;

<sup>5)</sup> H. Kemper, E.T.Z. 59, 391-395. 1938.

magnetic permeability smaller than unity, as has been demonstrated by Braunbek<sup>6)</sup>. Returning to the arrangement of fig. 1, it is possible to stabilize the equilibrium of the small permanent magnet  $M_2$  by placing a diamagnetic body  $G$  (fig. 2) closely below it. The latter exercises an upward force of repulsion upon  $M_2$ . Under suitable conditions (e.g. if  $M_2$  is sufficiently small compared to  $M_1$ ) any slight lowering of  $M_1$  from the state of equilibrium results in an increase in the repulsion exercised by  $G$  that is greater than the decrease in the attraction by  $M_2$ :  $M_1$  hence returns to the position of equilibrium. A slight upward displacement of  $M_1$  is compensated in an analogous way. As regards horizontal displacements of  $M_2$ , the equilibrium is stable in the absence of the diamagnetic body, and this situation is not altered by introducing it, provided that its horizontal extent is sufficiently large.

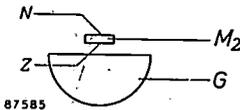
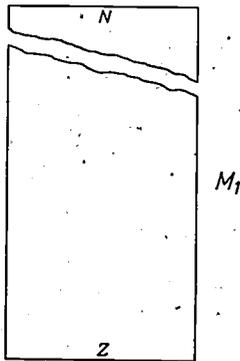


Fig. 2

We made an experimental model of the arrangement described here with  $M_1$  a "Ticonal" bar magnet 30 cm long and 3 cm in diameter, and a body  $G$  of very pure graphite, for which  $\mu_r = 0.9996$  (about the strongest diamagnetism attainable in solids, at normal temperatures.) The levitated magnet  $M_2$  was a cylindrical ferroxdure disc with a diameter of 1 mm and a thickness of 0.3 mm.

Other arrangements for producing levitation are possible when diamagnetic substances are used. Braunbek succeeded in levitating a piece of graphite weighing 75 milligrams between the specially shaped poles of an electromagnet, with a field strength between the poles of about  $1.8 \times 10^6$  A/m (23 000 oersteds). This extremely great field strength was required because the levitation is based on the homogeneity of the field, which must be greater according as  $\mu_r$  approaches unity and which has to extend over a distance comparable to the dimensions of the levitated body.

A field  $H$  exerts on a volume element  $dV$  of a substance with  $\mu_r < 1$  a force in the direction  $z$  given by

$$-\frac{1}{2} \mu_0 (1 - \mu_r) \frac{\partial H^2}{\partial z} dV$$

<sup>6)</sup> W. Braunbek, Z. Physik **112**, 753-763, 1939, also **112**, 764-769, 1939.

where  $\mu_0$  is the permeability of free space. If this force is to compensate the weight  $\rho g dV$  ( $\rho$  being the density,  $g$  the constant of gravitation), then we must have

$$\frac{\partial H^2}{\partial z} = -\frac{2g}{\mu_0} \frac{\rho}{1 - \mu_r} \dots \dots \dots (1)$$

For graphite,  $\rho/(1 - \mu_r)$  is  $2.3 \times 10^7$  kg/m<sup>3</sup>; for most other diamagnetic substances it is far greater. For the levitation of graphite the field must therefore obey the relation

$$\frac{\partial H^2}{\partial z} = 3.6 \times 10^{14} \text{ A}^2/\text{m}^3 (\approx 5.7 \times 10^8 \text{ Oe}^2/\text{cm}).$$

The very large field strength required by Braunbek could be obtained only with the aid of an electromagnet, so that he did in fact require energy for this levitation. The power required was very great, viz. 7 MW/kg. A dimensional analysis of the problem shows that for a given configuration, the maximum value of  $H$  required to obtain a given inhomogeneity ( $\partial H^2/\partial z$ ) decreases with the square root of the dimensions. In this way it proved possible to repeat Braunbek's experiments in a smaller scale, using a permanent magnet, as shown in fig. 3.

Apart from diamagnetic materials there is another group of substances with  $\mu_r < 1$ , viz. superconductors, for which  $\mu_r = 0$ . A far smaller degree of inhomogeneity of the levitating field is therefore adequate (cf. eq. 1), so that it would be possible in an arrangement similar to that of fig. 3 to levitate a comparatively large superconducting body by means of a far weaker magnet.

Using superconductors, the latter arrangement may be reversed to accomplish a spectacular result: a permanent magnet can be floated freely above a cup-shaped superconducting body<sup>7)</sup>. This experiment may be elucidated as follows. Consider first the case of two ferroxdure discs, one placed above the other, magnetized in the direction of their axes and repelling each other (fig. 4a). If the lower disc is held steady at a certain height, the top one is in a state of stable equilibrium as regards vertical displacement; however it can never at the same time be stable for horizontal displacement as well (Earnshaw's theorem). Now let us imagine the lower magnetic disc being removed and the space below the former plane of symmetry being occupied by a superconductor (fig. 4b). This will act upon the remaining magnetized disc with exactly the same repulsive force as before, since the magnetic field

<sup>7)</sup> The experiment, in which a magnet weighing about 1 g was used and cup and magnet were placed in a Dewar flask above liquid helium, has been described by V. Arkadiev, Nature **160**, 2, 330, 1947. The cup shape is used to give stable instead of a neutral equilibrium in the horizontal direction.

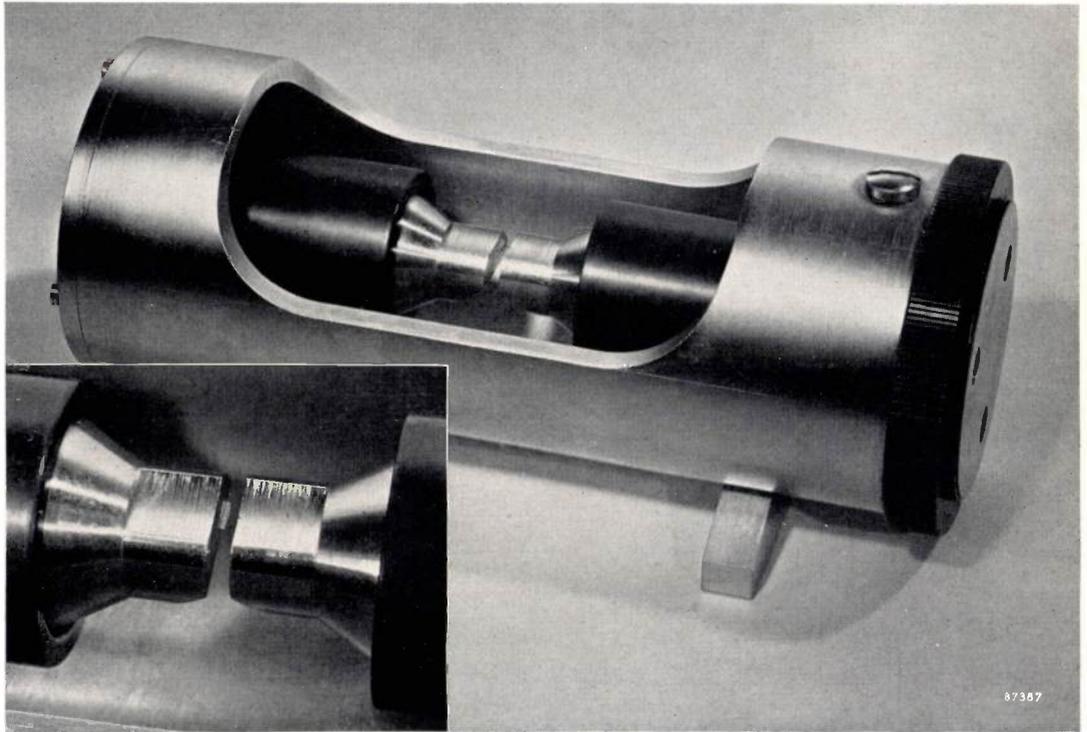


Fig. 3.

above the superconductor can be expressed as the sum of the field of the remaining magnet and that of its *mirror-image* in the surface of the superconductor<sup>8)</sup>. The magnet thus remains poised above the superconductor, but the difference from the case

of fig. 4a is that now the mirror image remains at all times vertically below the magnet, thus obviating the instability of the equilibrium for horizontal displacement.

If we finally imagine a diamagnetic substance being substituted for the superconductor, the same argument remains valid; only the magnetic strength of the mirror image will be reduced by a factor of  $(1 + \mu_r)/(1 - \mu_r)$ , which means by a factor of  $2 \times 10^4$  for graphite. Simple dimensional analysis demonstrates, however, that this far weaker repelling force will still be adequate to compensate the weight of the magnet, if only the latter is small enough. A permanent-magnet particle of suitable shape and with dimensions of a few microns should be capable of floating freely at a very low height, likewise of the order of a few microns, above the surface of a diamagnetic substance such as graphite.

A. H. BOERDIJK.

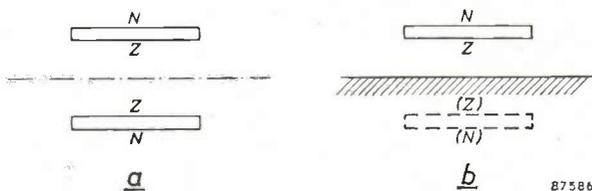


Fig. 4.

<sup>8)</sup> For completeness, it should be pointed out that substitution of a substance with  $\mu_r = \infty$  (ideal ferromagnetic) for the superconductor would be equivalent to creating a mirror image of the object magnet with *reversed* polarity, i.e. with its magnetization in the same direction as that of the object magnet.

## COLD-CATHODE TRIGGER TUBES

by C. H. TOSSWILL \*).

621.373.444:621.387

*The triode cold-cathode glow-discharge tube made its first appearance about twenty years ago as an alternative to the hot-cathode thyatron: it consumed no stand-by power and was immune to heater or filament breakage, but could only be used where the current demand was small and where the high maintaining potential and slow decay of the cold-cathode discharge were acceptable. These triodes have found many applications, notably as selectors in remote party-line telephone-switching; as counting elements in rings and other arrays; and as information storage devices, where their irreversible triggering action has been exploited. This article describes some of the improvements made in recent years in the design of cold-cathode trigger tubes.*

### Introduction

The work to be described here began in response to certain criticisms of existing cold-cathode tubes. It has had two distinct stages: first, the design of a general purpose experimental cold-cathode tube with more closely-defined and stable characteristics, and second, the development of two specialised forms suited to service in radiation monitors<sup>1)</sup>.

### Principle of operation

The triode cold-cathode tube is founded upon two fundamental properties of the low-pressure glow-discharge: first, the potential difference across a gap needed to initiate a glow-discharge is greater than that subsequently required to maintain it, and, second, the initiating or breakdown potential can be reduced by the insertion of a subsidiary or trigger discharge. Therefore, if a potential lying between the breakdown ( $V_b$ ) and maintaining ( $V_m$ ) values is applied to the main gap, no discharge will pass until a suitable trigger discharge is supplied. The main discharge will then continue until the voltage across the main gap is reduced to below the maintaining potential  $V_m$ . As the current carried by the main discharge may greatly exceed that carried in the trigger discharge, a power gain is available.

In a triode, the trigger discharge passes between one main electrode and a third, "trigger"; electrode, and has an effect dependent upon its duration and the tube geometry. The relation between minimum steady trigger current ( $I_t$ ) and main gap or anode potential ( $V_a$ ) shown in *fig. 1* applies to the case where both discharges have a common cathode, and is often known as the *transfer characteristic*. Certain

time delays, the statistical and formative lags (see below), affect the development of the trigger discharge itself. Always, however, the influence of the trigger is restricted to the beginning of the main discharge; after this, the trigger has no effect and the only means of interrupting the discharge is to lower the main gap potential beneath  $V_m$ . Such reduction

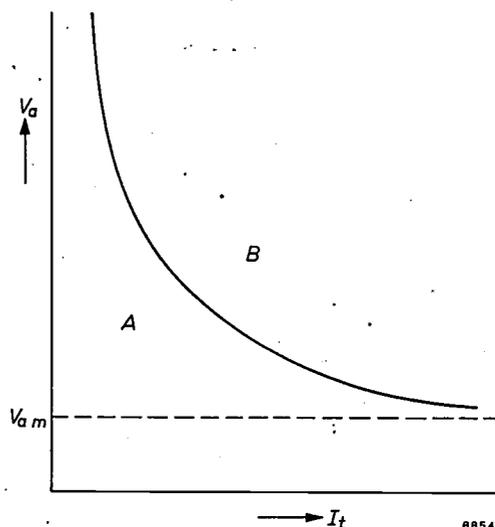


Fig. 1. Transfer characteristic of a trigger tube. This curve gives the minimum steady trigger current  $I_t$  necessary at a given anode voltage  $V_a$  to initiate breakdown. Once breakdown occurs (at the line dividing regions A and B) the tube continues to conduct so long as the anode voltage remains above the maintaining potential  $V_m$ .

causes the gas to revert to the condition prior to the passage of the discharge, to be "de-ionized", in a certain characteristic time. In the method of extinction and in the irreversible action of the control electrode the cold-cathode triode thus resembles the hot-cathode thyatron.

The breakdown characteristics of cold-cathode trigger tubes may be represented by a closed loop

\* ) The work described in this article was carried out at Mullard Radio Valve Co., Mitcham, Surrey.

<sup>1)</sup> E. Franklin and J. Hardwick, The design of portable gamma and beta radiation measuring instruments, *J. Brit. Inst. Radio Engrs.* **11**, 417-434, 1951. D. Taylor, Radiac instrumentation, *J. sci. Instr.* **29**, 315-322, 1952.

such as that shown in *fig. 2* which refers to a triode, the Mullard Z 300T (*fig. 3*). Points lying inside this loop correspond to anode and trigger voltages which do not cause a discharge. In order to fire the tube, the anode voltage and/or trigger voltage must change in such a way that the loop is crossed somewhere.

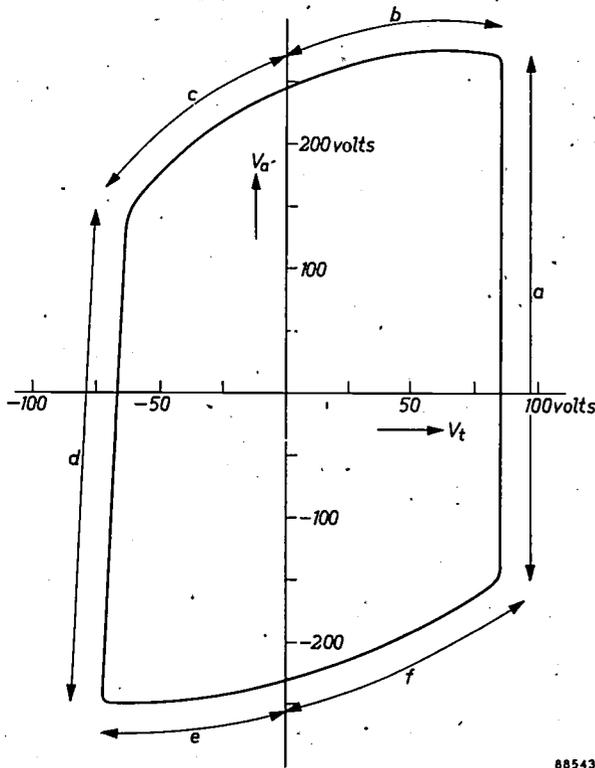


Fig. 2. Breakdown characteristic of Z 300T tube. For the tube to fire the closed loop must be crossed somewhere. The configuration of the trigger discharge depends on where the loop is crossed (regions a-f, see text).

Once the line has been crossed and the discharge initiated, the anode and trigger voltages may once more correspond to a point inside the loop. The section *a* of the loop refers to the firing of the tube by a discharge from trigger to cathode (direction of positive current flow). It is evident from the graph that the triggering voltage required is here independent of the anode voltage. A discharge from anode to cathode occurs when section *b* of the loop is crossed. Section *c* refers to a discharge from anode to trigger, section *d* from cathode to trigger, section *e* from cathode to anode and section *f* from trigger to anode.

The barium oxide layer on the cathode of the Z300T tube gives it a work function lower than that of the anode, so that the tube has rectifying properties. For given supply voltage and series resistance the current flow from anode to cathode and trigger to cathode is therefore larger than in the reverse direction. For this reason, this tube is operated in

quadrant 1 of *fig. 2*, where anode and trigger voltages are positive.

#### *Mechanism of breakdown; statistical and formative lags*

When a potential difference of a few volts is applied between cold electrodes in a gas, the latter behaves substantially as a perfect insulator<sup>2</sup>). Ionization due to cosmic radiation will give rise to a very small current, which will increase to a saturation value as the field is increased to a point where all the free electrons and ions are drawn to the electrodes. This saturation current is not constant with time, however, being subject to wide statistical variations owing to the random nature of the ionization. The saturation current will be greater and largely free of fluctuations if the cathode is exposed to light so as to emit a copious supply of photo-electrons.

As the applied field is increased, the free electrons become capable of imparting energy to the outer electrons of the gas atoms: the latter may be excited in the normal way and emit radiation, or they may be raised to metastable states, or, when the colliding electron has sufficient energy, some atoms may be ionized. With the gas pressures and electrode separation commonly used in trigger tubes, the mean free path of the free electrons will be only a fraction of the inter-electrode distance: hence ionization by collision will occur only when the applied potential is several times the ionization potential of the gas.

The positive ions in the gas acquire much the same energy as the free electrons but, because their mass is equal to that of the gas atoms, they give up on an average about half their kinetic energy at each collision. Since, also, elastic collisions are more probable than in the electron

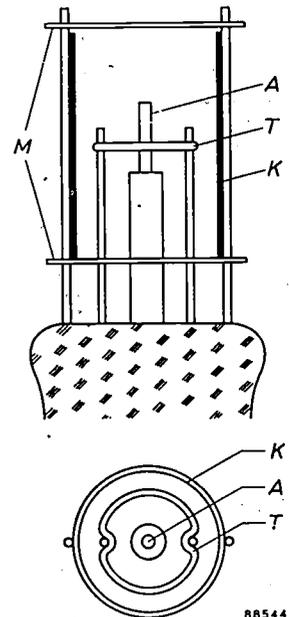


Fig. 3. Electrode structure (schematic) of the Z 300T cold-cathode tube. *A* anode, *T* trigger electrode, *K* cathode, *M* mica spacers. The tube is filled with argon to a pressure of 8 mm Hg. The electrodes are of nickel, the cathode having a coating of barium oxide upon its inner face.

<sup>2</sup>) For a general review of the field of gas discharge phenomena see F. M. Penning and M. J. Druyvesteyn, *Rev. mod. Phys.* 12, 1940.

case and charge-exchange may occur, the positive ions contribute little to ionization by collision.

With the onset of ionization due to electron collision, the anode current increases rapidly with the applied potential. If the number of new free electrons created by a single electron moving through unit potential is  $\eta$  (the gas ionization coefficient), the number of new free electrons created by  $n$  electrons traversing an element  $dl$  between the electrodes is

$$dn = n\eta E dl.$$

Assuming that  $n_0$  electrons leave the region of the cathode per second and that the field  $E$  is uniform, the number of electrons arriving at the anode per second is:

$$n_a = n_0 e^{\eta EL} = n_0 e^{\eta V},$$

where  $L$  is the electrode separation and  $V$  the applied potential.

The anode current may then be written

$$I = I_0 e^{\eta V}, \dots \dots \dots (1)$$

$I_0$  being the cathode emission current corresponding to some electron-supplying mechanism, e.g. photo-emission.

So long as the anode current is determined by (1) the discharge cannot become self-sustaining: if  $I_0$  ceases, (by shutting off the light in the case of photo-emission) the anode current will vanish. Physically, this means that when the free electrons initially present and those created by ionization have been drawn to the anode, the anode current must cease unless further free electrons are supplied. Hence, for the development of a self-sustaining discharge a secondary electron-supplying process is required. This may be effected by the positive ions; which on striking the cathode, give rise to further electron emission which enables the discharge to become self-sustaining. The number of new cathode electrons emitted per ion-pair formed between the electrodes is termed the secondary ionization coefficient  $\gamma$ . It is closely related to the work function  $\phi$  of the cathode; for a given gas,  $\gamma$  is larger the smaller  $\phi$ .

Taking this secondary process into account, the relation between  $I$  and  $I_0$  becomes:

$$I = I_0 \frac{e^{\eta V}}{1 - \gamma[e^{\eta V} - 1]} \dots \dots (2)$$

The characteristics represented by this expression are plotted in fig. 4 for various values of the initial cathode current  $I_0$ .

When the voltage  $V$  across the electrodes has the value  $V_b$  defined by

$$\gamma[e^{\eta V_b} - 1] = 1, \dots \dots (3)$$

the conclusion from (1) that  $I \rightarrow 0$  as  $I_0 \rightarrow 0$  is no longer necessarily valid, for equation (2) then leaves  $I$  indeterminate. What actually happens can be seen from fig. 4. The voltage  $V$  across the tube and current  $I$  are determined for a given value of  $I_0$  by the intersection of the relevant curve with the load line  $R$  (curved owing to the logarithmic scale)

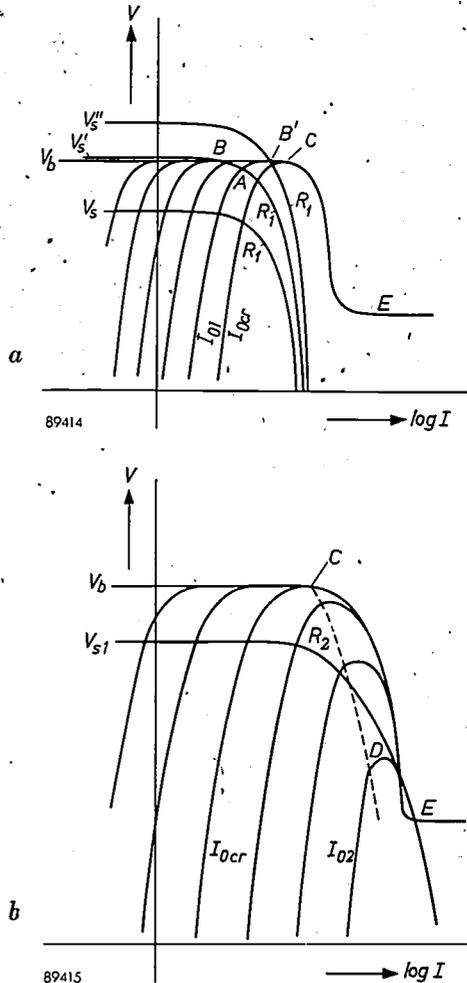


Fig. 4. Voltage-current characteristics (schematic) for a discharge between two electrodes in a gas, for various values of the initial current  $I_0$ . For clarity the current is plotted logarithmically. The numerical value of  $I_0$  for each curve is given by its intercept on the current axis.  
 a) Normal self-breakdown, with  $I_0 < I_{0cr}$ . This refers to breakdown in the trigger gap.  $R_1$  are load lines (curved owing to log scale) corresponding to supply voltages  $V_s, V_s'$  and  $V_s''$ . When the supply voltage exceeds  $V_b$  (e.g.  $V_s''$ ) breakdown occurs and the current which then flows for an initial current of  $I_{01}$  (say) is given by the point A. If  $I_0$  then  $\rightarrow 0$ , the discharge stabilizes at B. For a higher supply voltage ( $V_s'$ ) the discharge stabilizes at B'. The trigger discharge may also burn in the glow-discharge region E if the series resistor  $R_1$  is small enough.  
 b) Triggered breakdown ( $I_0 > I_{0cr}$ ) of the anode gap. When  $I_0 > I_{0cr}$ , space charge effects begin to appear which cause the breakdown voltage to be lowered. The dotted curve from the point C represents the locus of the points at which space charge begins to have an effect. For an initial current of  $I_{02}$ , the breakdown voltage is lowered to the point D.  $R_2$  is the load line corresponding to a supply voltage  $V_{s1}$  and series resistance  $R_2$  and is tangential to the curve  $I_{02}$ . In the absence of  $I_{02}$  the tube will be non-conducting but when  $I_0$  is raised to the value  $I_{02}$  by the trigger discharge, breakdown occurs and the tube current stabilizes at the point E.

corresponding to the supply voltage  $V_s$  and the resistance in series with the source.

As the supply voltage  $V_s$  is raised (fig. 4a), the load line is displaced parallel to the ordinate and at a voltage somewhat higher than  $V_b$  a self-sustaining discharge becomes possible; the voltage and current between the electrodes are then represented by a point such as  $A$  (fig. 4a). The discharge is self-sustaining for, as  $I_0 \rightarrow 0$  the working point on the load line moves to the left until intersection occurs at a voltage  $V = V_b$  (e.g. point  $B$  in fig. 4a).

With further increase in  $V_s$  the current  $I$  increases ( $B \rightarrow B'$ ) but  $V$  remains at  $V_b$ . At a certain value of  $I$  (point  $C$ ) space charge effects begin to distort the hitherto uniform field; further increase in  $I$  is accompanied by a fall in the voltage  $V$  across the electrodes and by the development of the glow discharge (region  $E$  in fig. 4a).

For values of the initial cathode current  $I_0$  less than a certain critical value  $I_{0cr}$ , the breakdown voltage (i.e. the maximum voltage  $V_{max}$  reached in the establishment of a self-sustaining discharge) has a constant value  $V_b$  determined solely by the materials, geometry and gas filling of the tube. Above this critical value, however, the breakdown voltage  $V_{max}$  becomes a function of  $I_0$ . Space charge effects now appear at an early stage and prevent  $V$  from reaching  $V_b$ . Consider the characteristic  $I_{02}$  ( $> I_{0cr}$ ) and a series resistor  $R_2$  giving the load line  $R_2$  in fig. 4b. As the supply voltage  $V_s$  is increased to the value  $V_{s1}$ ,  $V$  passes its maximum value near  $D$  and breakdown occurs. The voltage  $V$  across the electrodes drops sharply and, for a not too large value of the load (i.e.  $\leq R_2$  in this case), current and voltage stabilize at a point such as  $E$  corresponding to the glow discharge regime.

Normal *self-breakdown* between two electrodes in a gas (fig. 4a) is the process taking place in the *trigger gap* of a cold-cathode tube.  $I_0$  is normally much less than  $I_{0cr}$ . A pulse with a voltage greater than that corresponding to the gap breakdown voltage  $V_b$  (hereafter designated  $V_{tb}$ ) causes a self-sustaining discharge to be set up in the trigger gap for the duration of the trigger pulse. (A condenser is often connected between trigger and cathode to intensify or prolong the trigger discharge.) A fraction of the charge carriers of this discharge are transported to the main anode-cathode gap and form the " $I_0$ " to initiate the main discharge, as outlined below. (Whether the trigger discharge burns in the Townsend region — the current region up to the point  $C$ , fig. 4 — or in the glow discharge region near  $E$ , depends on the circuit conditions. These must, of course, be so chosen that the trigger discharge is of

sufficient intensity to provide the " $I_0$ " necessary to ignite the main gap — see below.)

*Triggered breakdown* is the process by which the *main anode-cathode* discharge is started. The voltage at which a self-sustaining discharge across this gap can start is temporarily lowered by providing an initial current  $I_0$  that exceeds the value of  $I_{0cr}$  for this gap (fig. 4b). As mentioned above, this  $I_0$  is derived from the already-established trigger discharge, which can easily be made to supply charge carriers at such a rate that  $I_{0cr}$  is exceeded. (The lowering of anode breakdown voltage as a function of trigger current  $I_t$  was shown schematically in fig. 1). If, then, across the main gap, a steady potential is applied somewhat below the normal self-breakdown value  $V_b$  (hereafter designated  $V_{ab}$ ), the tube will remain quiescent until triggered by the presence of an adequate initial current  $I_0$  due to a discharge set up in the trigger gap.

There is always a certain delay after the application of the trigger pulse before the self-sustaining discharge is set up in the trigger gap. The delay is made up of two components. The first, which is by no means constant, is known as the *statistical lag* and arises from the necessity for the presence of at least one favourably-placed electron to enable the discharge to start. In practice, if  $I_0 \ll I_{0cr}$  an uncertain and frequently lengthy delay occurs. With increase in  $I_0$  the statistical lag is gradually eliminated. It is desirable that the initial current  $I_0$  lies between the value just able to remove the statistical lag and the value  $I_{0cr}$ . The inter-electrode current which then flows is called a *priming current*. With many cathodes a suitable level of ambient illumination will give rise to such a priming current; other means for providing a priming current will be described below.

The other delay is the time interval required to build up the self-sustaining discharge from the moment the first favourably-placed electron appears until the discharge has reached its equilibrium value. This delay is called the *formative lag*; it does not vary randomly but it depends sharply on the excess voltage of the trigger pulse above the breakdown voltage  $V_b$  of the trigger gap. To keep this delay small, the trigger pulse voltage should exceed the trigger breakdown voltage by some tens of volts (e.g.  $V_s''$  in fig. 4a).

#### *Desirable characteristics of cold-cathode tubes*

To widen the field of application of trigger tubes the following should be aimed at:

- a) Good reproducibility of breakdown and maintaining potentials, both from tube to tube and in individual tubes during life.

- b) Faster operation, if possible approaching the response of hot-cathode devices.
- c) Improvement of input sensitivity.
- d) Reduction or elimination of random operating delays, even at low levels of ambient illumination.
- e) Reduction of incremental impedance and noise level; for some uses, e.g. for speech-channel applications, the noise level of most tubes is too high.
- f) Small dimensions.

It would not be possible to improve on all these points simultaneously, as some of them are partly contradictory. The initial aim, therefore, was to evolve a design and a manufacturing technique to give a cold-cathode tube in which some of the existing characteristics were improved, viz. points *a*, *d* and *f*: uniformity of characteristics and stability during life, freedom from random delays (statistical lags) and robust miniature construction. These particular properties were selected for improvement because it appeared that they were not incompatible and also because they are collectively of importance in many applications<sup>3)</sup>.

It will be noted that all three features are ingredients of reliability, the improvement of which was indeed the primary aim. It will further be noted that while the raising of the intrinsic input sensitivity was not considered at this stage, the effective sensitivity of the whole circuit is of course improved when the tube characteristics are accurately known and maintained, for the input signals may then safely be smaller without fear of the tube not responding or, worse, spurious breakdown when not signalled.

#### Improvement of reliability and performance

The breakdown and maintaining potentials of cold-cathode tubes are inversely related to the values of the gas ionization coefficient  $\eta$  and the secondary ionization coefficient  $\gamma$ . It seems obvious therefore to choose gas fillings and tube materials which yield large and stable values of both coefficients: in fact, the inert argon filling and barium-on-nickel cathode of the Z 300T tube were selected on this basis. However, the fragile nature of the barium layer (which is susceptible to change even with

normal discharges and is removed entirely by sputtering if temporarily overloaded) together with gaseous contaminants from the wall of the tube and elsewhere, lead to changes in  $\eta$  and  $\gamma$  which are observed by the user as the objectionable variations mentioned above. The steps taken to stabilize these quantities are outlined below.

Another matter requiring attention is the elimination of the statistical lag. As we have seen, to do this, a priming current must flow in the otherwise dormant tube.

#### Molybdenum sputtering technique

In 1946 Penning and others<sup>4)</sup> described means of stabilizing the basic glow discharge quantities for certain materials prior to the successful measurement of the "normal cathode fall" (the constant drop in potential near the cathode surface, which occurs when the current through the tube is between certain limits). The sustained cathode sputtering technique employed in this work had obvious application in any glow-discharge device dependent upon stability of the operating potentials, and was soon employed in the voltage reference tubes 85A1 and 85A2, where a molybdenum cathode is used in conjunction with a filling of neon and argon<sup>5)</sup>. Similar arguments clearly favoured the introduction of this system in the class of trigger-tube under consideration. There are the following objections to the molybdenum-sputtering technique. Working potentials are high compared with those of tubes using barium or, say, potassium cathodes. Photo-emission arising from normal ambient illumination is ineffective in supplying the priming current necessary for the elimination of the statistical lag. Lastly, the manufacturing technique is not without its difficulties and the price of the materials is a factor of some significance. However, the advantages greatly outweigh the objections.

The successful application of the cathode sputtering technique to trigger-tube manufacture demands an electrode structure quite different to that of Z 300T (fig. 3). Thus, in the first place, mica spacers must be abandoned. They act as barriers to the transport of sputtered material to the walls of the tube, and when coated with this material, act as a source of inter-electrode leakage. Also, mica is a laminar material, and often secretes contaminants which can be slowly released during the operation of the tube. Secondly, the cylindrical form of the

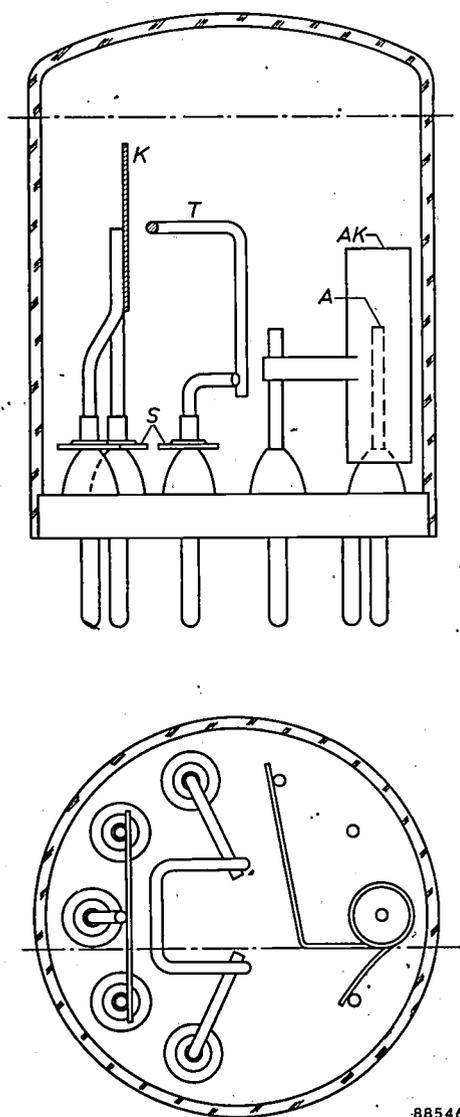
<sup>3)</sup> Others have aimed at improvements in directions not attempted in the present work. An experimental tube suitable for speech channel applications, the Z 500T, has been developed in the Philips Laboratories in Eindhoven; see J. Domburg and W. Six, Philips tech. Rev. 15, 265-280, 1953/54. Other work in this field is reported by M. A. Townsend and W. A. Depp, Bell. Syst. tech. J. 32, 1371-1391, 1953. Work on improving the speed of response is described by G. H. Hough and D. S. Ridler, Electronic Eng. 24, 152-157, 1952.

<sup>4)</sup> F. M. Penning and J. H. A. Moubis, Philips Res. Rep. 1, 119-128, 1946; T. Jurriaanse, F. M. Penning and J. H. A. Moubis, Philips Res. Rep. 1, 225-230, 1946; and T. Jurriaanse, Philips Res. Rep. 1, 407-418, 1946.

<sup>5)</sup> T. Jurriaanse, Philips tech. Rev. 8, 272, 1946.

cathode is not suited to sputtering. The cathode must have a shape which allows the unimpeded escape of sputtered material, and must be so placed that while the walls are covered with a uniform film, the lead-out area remains bare. Thirdly, the tube envelope must be small enough to be covered in a reasonable time, and also be free from re-entrant portions such as exist in the old "pinch" construction.

Experiments showed that a small oblong cathode plate mounted with its plane parallel to the axis of the tube, in a standard 7-pin or 9-pin miniature envelope, met all these needs very well. This structure is shown in *fig. 5*. Subsequent field experience has shown that this rather frail-looking structure is well able to withstand the shocks of normal service.



88546

Fig. 5. Electrode structure (simplified) of experimental Mo-sputtered tetrode trigger tube. *K* cathode, *T* trigger electrode, *AK* tubular auxiliary cathode surrounding the anode *A*. *SS* are shields which leave a region near the lead-out wires free of sputtered material (see below).

#### *Provision of priming current*

Methods of providing a priming current include:

- The admission of normal daylight — the simplest solution, and that used in Z 300T. However, this is ineffective with the molybdenum cathode, whose photo-emission is negligible for the spectral range transmitted by normal glasses. Moreover, reliance on the photo-electric effect imposes undesired restrictions upon the conditions of use of the tube.
- The use of radio-active materials. This would have been a more satisfactory solution, but was rejected because of lack of experience with "soft" emitters whose influence would not be objectionable outside the tube, and because the third system has additional advantages quite apart from the priming.
- The use of an auxiliary discharge. Such a discharge may be direct, passing the trigger and cathode, or indirect, achieving its effect by the transport of some of its ions into the trigger-cathode space, possibly over a considerable distance.

The direct auxiliary discharge imposes certain restrictions on the trigger circuit, but is adopted in one of the tubes described below (Z 801 U). An upper limit somewhat below 1 micro-ampere is set on the permissible value of such a direct priming current by the need to avoid distorting the applied fields within the tube by the accumulation of space-charge.

The indirect discharge, which may be larger, is used in the stabilizer tube Z 800 U — with the discharge flowing to the anode from an auxiliary cathode. The latter is of tubular form, and surrounds the wire anode (*fig. 5*). This tubular auxiliary cathode provides an economical solution to two problems: how to obtain within a miniature envelope an effective spacing large enough to create a useful difference between  $V_{ab}$  and  $V_{am}$ , and how to screen a discharge between anode and auxiliary cathode from the trigger-cathode space. By varying the recession of the anode tip within the auxiliary cathode,  $V_{ab}$  may be altered without affecting the remainder of the trigger-tube, while the presence of the auxiliary discharge stabilizes the potential of the auxiliary cathode relative to the anode. The auxiliary discharge may now reach 10  $\mu A$ , and so enter the current range controllable by standard resistors, without exerting an excessive influence on the field in the trigger-cathode gap.

With the introduction of the auxiliary discharge, the maximum operating delay is reduced from about 1 sec to the order of 100  $\mu$  sec, which is essentially the formative lag.

#### *Experimental trigger tube*

The experimental tube (*fig. 5*) incorporating the features outlined above showed satisfactory behavi-

our. In order to keep the triggering voltage  $V_{tb}$  to a minimum, the trigger-cathode separation was kept small (1-2 mm). In addition, a trace of argon (about 1%) was added to the gas filling of neon, at a pressure of 40 mm Hg. Under these conditions  $V_{tb} \approx 115$  V.

The recession of the anode tip within the tubular auxiliary cathode is effectively limited by the fact that, however much the recession, there is a limit to the attainable anode-cathode breakdown voltage ( $V_{ab}$ ) fixed by the breakdown between auxiliary cathode and cathode. Beyond a certain point, therefore, further recession will not enhance the value of  $V_{ab}$ . For the gas filling used,  $V_{ab}$  has its maximum at 250 V. Under these conditions, the discharge-maintaining voltage  $V_{am} = 100$  V.

The triggering voltage stability of early experimental tubes was not entirely up to expectation. This was found to be due to contamination of the trigger electrode, which was then of nickel. So long as different materials were used for trigger and cathode, it proved impracticable to clean both by sputtering. The use of a molybdenum trigger gives a satisfactory stability, but it has the disadvantage that in circuits in which the trigger may become earthed, the maximum operating voltage may have to be smaller to preclude spurious trigger-anode breakdown.

Variations in the discharge-maintaining potential  $V_{am}$  are similar to those of the voltage reference tube (diode) 85A2 and are satisfactory.

#### Self-quenching operation of trigger tubes

The achievement of better stability and the virtual elimination of statistical lags cleared the way for the design of two trigger tubes for certain specific applications, viz. for use in a stabilizer circuit (Z 800U) and for use in a rate meter circuit (Z 801U)<sup>6)</sup>. The former has to respond to very small continuous currents and the latter to very small pulses of current.

While both the Z 800U and the Z 801U may be used in certain conventional cold-cathode tube circuits, they are primarily designed for the oscillatory self-quenching circuit. In this type of circuit (fig. 6), the anode current is derived from the discharge of a condenser and is therefore intermittent, passing as a single pulse each time the tube is triggered. The self-quenching circuit has the advantage that the trigger electrode is capable of exercising a quasi-continuous and reversible control over the mean anode current. This gives the circuit a number of important applications. Trigger tubes working in such circuits face conditions differing markedly from those prevailing in steady current circuits. Of

<sup>6)</sup> These two special-purpose tubes are not available commercially from the Philips Electronic Tube Division.

particular importance are the factors governing extinction of the discharge after each current pulse. The self-quenching action is obtained with an anode load  $R_a$  which is considerably less than might be suggested from the fact that to extinguish an

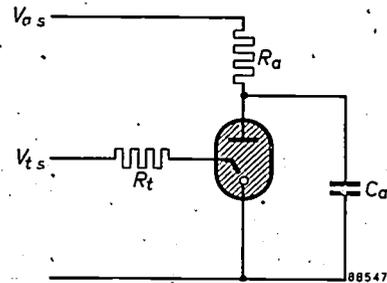


Fig. 6. Trigger tube in self-quenching circuit.

established *steady* discharge (the supply voltage remaining constant)  $R_a$  would have to be increased to at least  $10^8 \Omega$ . The self-quenching circuit can operate with  $R_a$  only a hundredth of this value.

A study of the self-quenching discharge leads to the following explanation. The absence of a resistor limiting the condenser discharge enables anode breakdown to be followed by an exceptionally large tube current; this has passed its peak and largely decayed by the time that  $C_a$  has discharged sufficiently for  $V_a$  to fall to the normal maintaining voltage  $V_{am}$ , and has, in consequence, created an abnormal positive ion cloud around the anode. As this cloud moves away, the anode is able to descend well below  $V_{am}$  in an "overswing"; at the end of the overswing the flow of current through  $R_a$  carries  $V_a$  up again to  $V_{a,m}$ , but not before the process of de-ionization has gone far enough to lead to the collapse of the discharge. The overswing, which gives the tube the appearance of having inductive properties, commonly carries  $V_a$  down to  $\frac{3}{4}$  of  $V_{am}$ . In the case of the Z 800U tube other considerations led to a design having satisfactory self-quenching properties; with the Z 801U, however, the main problem was the reconciliation of the self-quenching requirements with the remainder of the specification.

#### Stabilizer trigger tube

The trigger tube Z 800U<sup>7)</sup> (fig. 7) was originally developed for use in a stabilized H.T. supply for Geiger counters in radiation monitors<sup>1)</sup>, viz. the low-current stabilizer circuit introduced by Goulding<sup>8)</sup> (fig. 8). The tube has to meet the following three conditions:

- 1)  $V_{ab}$  to be approximately 300 V.
- 2) The tube to operate in a self-quenching circuit.
- 3) The trigger input current demand to be less than  $10^{-7}$  A.

<sup>7)</sup> C. H. Tosswill, British Patent No. 709 103.

<sup>8)</sup> F. S. Goulding, British Patent No. 732 776; and, A variable voltage stabilizer employing a cold-cathode triode, *Electronic Eng.* 24, 493-497, 1952 (the Z 800U is referred to as VX 8107, and the Z 801U as VX 8086); G. O. Crowther, Cold-cathode voltage stabilizer, *Electronic Eng.* 25, 127, 1953.

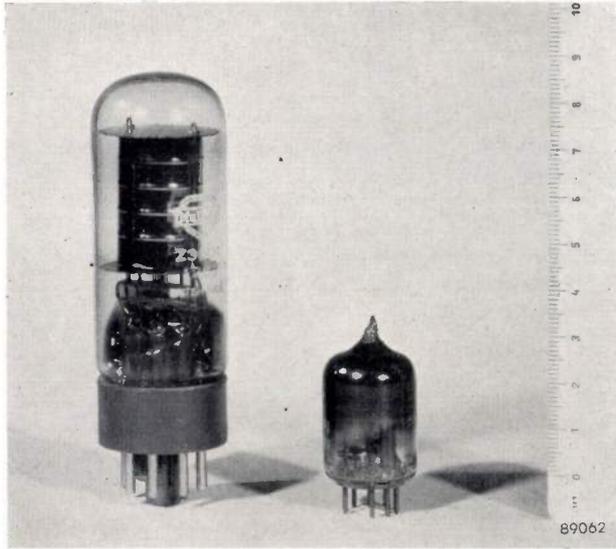


Fig. 7. Photograph of the Z 800 U trigger tube (centre, see also fig. 5) showing its size in relation to the Z 300 T tube mentioned earlier (fig. 3). (The scale on the right is in cm.) The tubular auxiliary cathode and the sputtering shields round the lead-in wires can just be seen through the envelope of the Z 800 U. The ratemeter trigger tube Z 801 U described later (see fig. 14) and the general purpose tube Z 803 U are identical in size and appearance.

With the same electrode structure as that used in the experimental tube (fig. 5), the required breakdown voltage was obtained by using a gas filling of neon with 3% argon at a pressure of about 40 mm Hg. This percentage of argon is somewhat greater than that used in the experimental tube (1%). The change also raises the maintaining voltage  $V_{am}$  to 105 V and the triggering voltage  $V_{tb}$  to 125 V.

The tube has to operate under circuit conditions which create self-quenching discharges in both the trigger-cathode and the anode-cathode gaps (fig. 9a, in which the priming electrode is omitted for simplicity). Both anode and trigger currents therefore

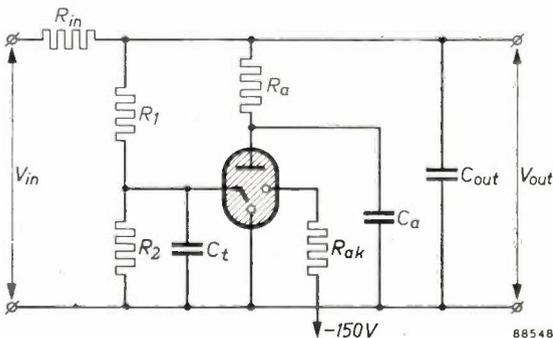


Fig. 8. Goulding's stabilizer circuit<sup>8)</sup> using the trigger tube Z 800U.  $R_1 = 68 \text{ M}\Omega$ ,  $R_2 = 100 \text{ M}\Omega$ ,  $C_t = 270 \text{ pF}$ ,  $R_a = 1 \text{ M}\Omega$ ,  $C_a = 0.005 \text{ }\mu\text{F}$ ,  $R_{ak} = 100 \text{ M}\Omega$ ,  $C_{out} = 0.25 \text{ }\mu\text{F}$ . In this stabilizing circuit a fraction  $R_2/(R_1 + R_2)$  of  $V_{out}$  is applied to the trigger electrode. Therefore if  $V_{out} > V_{tb}(R_1 + R_2)/R_2$  the tube will be triggered and the pulse train through the tube will cause  $C_{out}$  to discharge until  $V_{out}$  has fallen sufficiently to dispose of the inequality. Any further fall in  $V_{out}$  will cut off the tube current because the discharges are self-quenching. In this way a stabilizing action is secured (provided that  $I_{in} > I_{out}$ ) with  $V_{out}$  slightly in excess of  $V_{tb}(R_1 + R_2)/R_2$ .

take the form of trains of pulses and the interaction between them complicates the general analysis of the problem. We may confine ourselves to the important special case when the anode supply voltage  $V_{as} < V_{a,b}$  and  $R_a C_a \ll R_t C_t$ . Each anode pulse is then necessarily the immediate consequence of a trigger pulse, though individual anode waveforms are independent of the actuating trigger discharges (fig. 9b).

The mean anode current bears a constant relation to the mean trigger current, since both are determined by condensers charged to the (fixed) supply potential  $V_{as}$  and to the trigger breakdown potential  $V_{tb}$ . Since the anode current pulse train can be stopped as well as started by the action of the trigger, i.e. by variation of the trigger supply potential  $V_{ts}$ , the mean anode current  $I_a$  can be continuously controlled over a considerable range by the trigger supply voltage  $V_{ts}$ . The circuit of fig. 9a may therefore be used as a D.C. amplifier. Its amplification factor is  $Q_a/Q_t$ , i.e. the ratio of the charges on the anode and trigger capacitors.

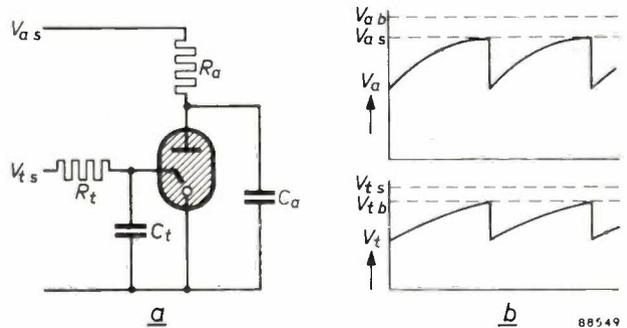
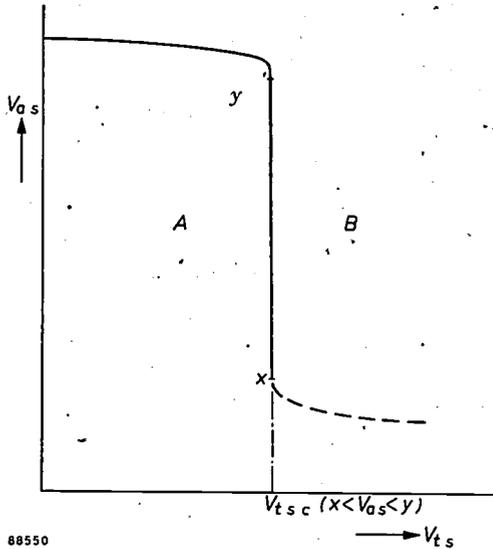


Fig. 9. a) Circuit giving self-quenching discharges in both anode-cathode and trigger-cathode gaps. Priming electrode (auxiliary cathode) omitted for simplicity.  $R_t = 100 \text{ M}\Omega$ ,  $C_t = 470 \text{ pF}$ ,  $R_a = 1 \text{ M}\Omega$  and  $C_a = 2000 \text{ pF}$ . b) Anode and trigger waveforms for the circuit shown in (a). When  $R_a C_a < R_t C_t$ , a trigger pulse above a certain amplitude gives rise to an immediate anode pulse. The anode waveforms, however, are independent of the form of the actuating trigger discharge. (The breakdown and extinction processes occupy an interval that is short compared with  $R_a C_a$  and  $R_t C_t$ , and therefore appear in the graph as a vertical line.)

Operation under the above conditions implies that both discharges are self-quenching (which can be achieved without using the very high component values which would be necessary under steady operating conditions) and that the charge passing in a single trigger pulse must be sufficient to cause anode breakdown. This requirement is analogous to the  $I_t/V_a$  relationship (fig. 1) for triggering in the case of the steady current circuit. In the self-quenching case one of the variables is thus a component value ( $C_t$ ); however it is more fruitful to consider another approach, viz. the effect of the trigger supply potential. (The variation of  $V_{ab}$  with  $C_t$  can be deduced from fig. 12.)

### Critical trigger supply potential ( $V_{tsc}$ )

Fig. 10 shows the anode supply potential  $V_{as}$  plotted against the trigger supply potential  $V_{ts}$ . The curve represents the conditions under which a flow of mean anode current  $\bar{I}_a$  can just be detected. The value of  $V_{ts}$  corresponding to the nearly vertical part XY is called the critical trigger supply



88550

Fig. 10.  $V_{as}$  plotted against the value of  $V_{ts}$  required to just give a mean anode current, for a self-quenching circuit. Within a considerable range of anode supply voltages (X-Y) the required trigger supply potential is independent of  $V_{as}$ . This value is termed the critical trigger supply potential,  $V_{tsc}$ .

potential; the tube is preferably operated in this region since  $V_{ts}$  is here nearly independent of the anode supply voltage  $V_{as}$ .

Further penetration of region B (fig. 10) leads to a growth of  $\bar{I}_a$ ; on return to region A,  $\bar{I}_a$  is cut off. This provides an interesting comparison with the irreversible action of fig. 2, where the current can assume only a certain fixed value.

The critical trigger supply potential  $V_{tsc}$  can be resolved into three components:

a) The breakdown voltage  $V_{tb}$  necessary across the actual trigger-cathode space to initiate breakdown. This has been discussed in the previous sub-section.

b) The potential difference developed across  $R_t$  (fig. 9a) due to ohmic leakage over the tube surfaces. This has been reduced to negligible proportions by reducing the leakage current to  $10^{-10}$  A. Two simple measures were sufficient to achieve this: first, the mounting of small shields upon each lead-out wire close to the point of entry into the glass, so as to create a small bare zone within which no sputtered material could fall; second the spraying of the exterior of the envelope in the neighbourhood of the lead-out wires with a silicone solution to inhibit the formation of conducting films of moisture.

c) The potential difference developed across  $R_t$  due to the pre-breakdown trigger-discharge current  $I_{tb}$  (the trigger condenser does not contribute to this current at this early stage). The maximum value of  $I_{tb}$  determines the current drain on the trigger supply which, as stated, must not exceed  $10^{-7}$  A.  $I_{tb}$  is dependent on a number of circuit and tube parameters; these will now be discussed.

### Trigger supply current drain at breakdown ( $I_{tb}$ )

The influence on  $I_{tb}$  of four variables was examined as outlined below.

1) Effect of anode supply potential  $V_{as}$ . Fig. 11 shows the effect of  $R_t$  on  $V_{tsc}$ . Since the changes in applied potentials were slow, the separation of the two curves can only be due to the voltage drop across  $R_t$ , due to surface leakage current together with  $I_{tb}$ . The leakage current ( $10^{-10}$  A) accounts for a separation of only 0.006 V for the two resistances used (10 M $\Omega$  and 68 M $\Omega$ ). Hence the observed separation of approximately 1 volt must be almost entirely due to  $I_{tb}$ ; the value of  $I_{tb}$  is thus  $10^{-10} \times 1/0.006$ , i.e.  $2 \times 10^{-8}$  A. It will be noted that the separation of the two curves is constant within a considerable range of anode supply voltages: this implies that  $I_{tb}$  is independent of  $V_{as}$ . Fig. 11 also shows that the critical trigger supply voltage is substantially independent of  $V_{as}$  (cf. fig. 2 for the steady discharge case).

2) Effect of auxiliary cathode current  $I_{ak}$ . Fig. 11 is based on a value of  $I_{ak} = 3 \mu\text{A}$ . Little change in

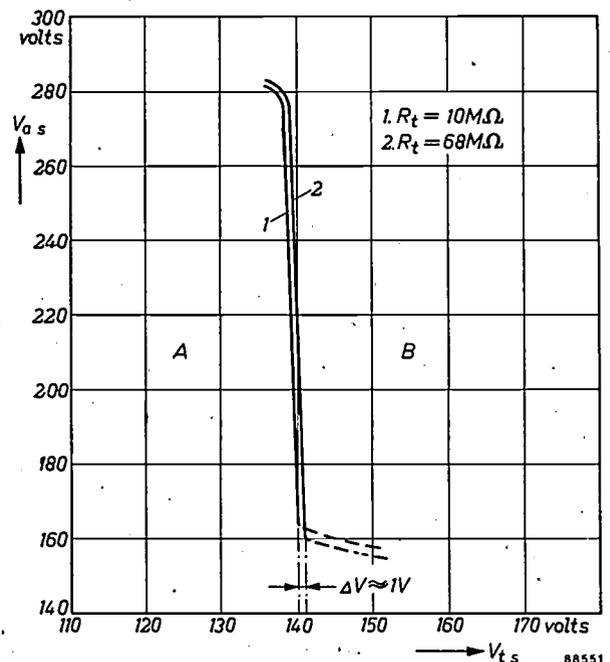


Fig. 11. Variation of  $V_{tsc}$  with  $V_{as}$  for two different values of the trigger resistor  $R_t$ .  $C_t = 1000$  pF. From the separation of the two curves the trigger supply current drain at breakdown can be deduced (see text).

$I_{tb}$  can be detected for variations of  $I_{ak}$  between 1 and 10  $\mu\text{A}$ . This somewhat unexpected result is probably due to the fact that the auxiliary discharge starts near the end of the cylindrical auxiliary cathode and expands inward with growth.

3) Effect of trigger condenser  $C_t$ . Fig. 12 shows again the variation of  $V_{tsc}$  with  $V_{as}$  but with  $R_t$  kept

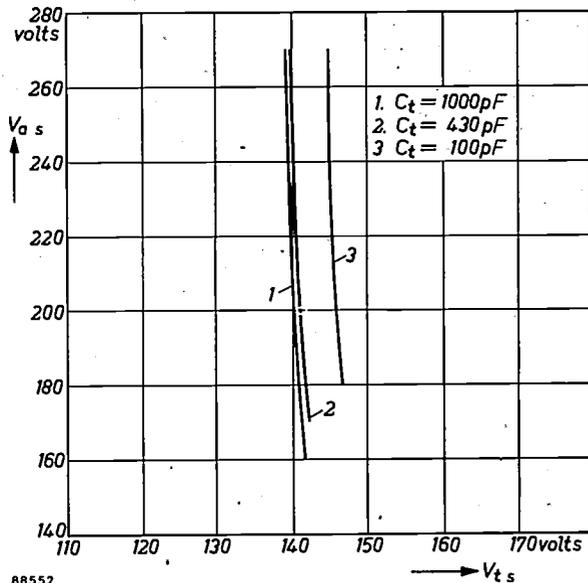


Fig. 12. Effect of trigger capacitance  $C_t$  on the  $V_{as} - V_{ts}$  characteristic. Trigger resistor  $R_t = 68 \text{ M}\Omega$ .

constant and three different values of  $C_t$ . The lower the value of the latter, the higher is the critical trigger supply voltage: this implies that  $I_{tb}$  is greater. Hence there is a lower limit to the capacitance of  $C_t$ , if  $I_{tb}$  is to be kept below  $10^{-7} \text{ A}$ .

4) Effect of trigger-cathode spacing. In the absence of mica spacers, a certain spread in the electrode spacings is inevitable. In tubes with smaller spacings large values of  $I_{tb}$  ( $10^{-6} \text{ A}$ ) were liable to occur intermittently, an effect visible as a corona surrounding the trigger wire instead of the normal glow adjacent to the cathode. This trouble was eliminated by increasing the trigger-cathode separation from 1 mm to 3 mm, whereby the trigger breakdown voltage  $V_{tb}$  is also increased, from 125 to 135 V.

#### Performance in stabilizer circuit

The investigations described above showed that only very minor modifications were necessary to the original experimental tube in order to get satisfactory operation in a self-quenching circuit with a very small trigger current drain.

In the restricted mode of operation defined earlier ( $V_{as} < V_{ab}$ ,  $R_a C_a \ll R_t C_t$ ), increase of  $V_{ts}$  above  $V_{tsc}$  gives rise to a finite mean anode current  $\bar{I}_a$  whose value for a given set of supply potentials

and component values depends on  $V_{tsc}$ , the anode extinction potential and the trigger extinction potential. Since  $V_{tsc}$  and the two latter quantities are constant for a given tube, the tube can be used for stabilizing a supply potential as in the Goulding circuit (fig. 8). The stabilizing action of the circuit will be clear from the figure and the details given in the caption.

This stabilizer, using the Z 800U tube, has the following important properties.

a) It works in a voltage range (200-300 V) which cannot be easily accommodated by either the glow discharge or the corona discharge diode.

b) Its current range is 5-150  $\mu\text{A}$ . Diode glow discharge tubes do not normally work successfully in this range. The corona stabilizer possesses a similar current range but cannot compete in life stability with the trigger tube circuit embodying the Z 800U.

c) The output voltage can be adjusted, for each tube, by altering the resistor ratio  $R_2/(R_1+R_2)$  (see fig. 8); the output voltage can in this way be varied over a considerable range to allow for the spread of Geiger-Müller tube characteristics.

The properties of the Z.800U tube may also be exploited in more conventional circuits in which the current flows continuously. Retention of the trigger condenser will still permit operation with small signal currents down to  $10^{-8} \text{ A}$ . If the auxiliary discharge be dispensed with and the occurrence of statistical lags accepted, operation with input currents down to  $10^{-10} \text{ A}$  is possible.

#### Rate meter trigger tube

The Z 801U<sup>9)</sup> was developed for use in rate meter circuits, that is to say, circuits whose output current is a measure of the rate of arrival of input pulses. The pulse source relevant to the development of this tube was a Geiger-Müller tube, but the results obtained are applicable to any system handling pulses of similar character.

The basic cold-cathode tube rate meter circuit devised by Franklin<sup>10)</sup> is shown in fig. 13a. The advantages of this circuit are described in full in the patent specification and will only be summarised here.

Each time the tube is triggered, the anode circuit goes through a single self-quenching cycle. Provided that the anode has sufficient time following the downward swing to regain the full  $V_{as}$  value before the arrival of the next input pulse, then  $\bar{I}_a$  will be proportional to the rate of triggering. The connection of the trigger to the anode through a very large

<sup>9)</sup> C. H. Tosswill, British Patent No. 696 077.

<sup>10)</sup> E. Franklin, British Patent No. 638 833.

resistance has the effect of poising the trigger close to the breakdown potential. Two purposes are served in this way: a direct priming current is provided for the trigger-cathode space without the need for a fourth electrode and even the smallest input

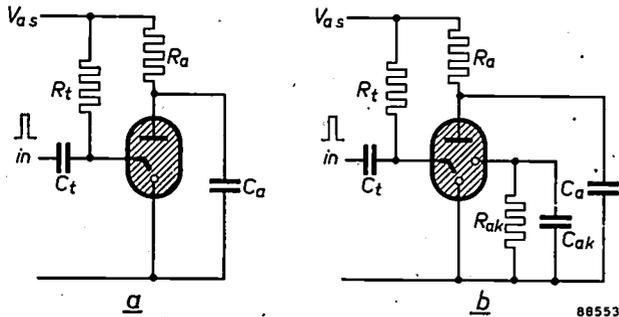


Fig. 13. a) Basic rate meter circuit according to Franklin<sup>10)</sup>. b) Franklin's rate meter circuit using a tetrode trigger tube. Typical component values:  $R_t = 100 \text{ M}\Omega$ ,  $C_t = 50 \text{ pF}$ ,  $R_a = 200 \text{ k}\Omega$ ,  $C_a = 0.01 \text{ }\mu\text{F}$ ,  $R_{ak} = 1 \text{ M}\Omega$ ,  $C_{ak} = 50 \text{ pF}$ .

pulse applied to the trigger will cause it to exceed the breakdown potential. For the successful working of this circuit three conditions must be fulfilled: a) The trigger priming current must not carry the trigger circuit to the point of instability, i.e. the priming current must be less than  $I_{tb}$ . Whether the trigger priming current  $>$  or  $< I_{tb}$  depends on the values of  $R_t$  and  $C_t$  (fig. 13a).

b) In this circuit, the input pulse will be able to deliver the whole of its charge ( $Q_{in}$ ) into the trigger-tube. For a given value of  $V_{as}$  there is a certain minimum value of  $Q_{in}$  for anode breakdown.

c) The anode and trigger discharges must be self-quenching.

An experimental triode was constructed to study the rate meter circuit requirements. The molybdenum-sputtering technique was again used and the same neon-argon filling was used as in the stabilizer tube. The leakage-counteracting screens of the stabilizer tube — essential to keep the leakage resistances high compared with  $R_t$  ( $R_t \approx 10^8 \text{ ohms}$ ) — were also fitted. In order to meet the radiation monitor requirements, it is additionally necessary that:

a) The "extinction current" should be at least  $210 \text{ }\mu\text{A}$  to give an adequate value of  $\bar{I}_a$  at the anticipated operating frequency. By the term "extinction current" is meant the current passing through  $R_a$  at the lower end of the anode swing, with  $R_a$  set at the minimum value compatible with self-quenching operation. The "extinction current" in the case of the triode was only  $70 \text{ }\mu\text{A}$ .

b) The input charge ( $Q_{in}$ ) necessary for reliable response to Geiger Müller tube signals, should be not more than about  $5 \times 10^{-11}$  coulomb with the anode

supply voltage  $V_{as}$  safely less than the breakdown voltage  $V_{ab}$ . The value for the experimental triode was  $5 \times 10^{-10}$  coulomb.

c) The frequency range should preferably be about  $10 \text{ kc/s}$ . The triode had a range of only  $1 \text{ kc/s}$ .

These aims are difficult to reconcile, improvement of one being at the expense of the others. However, by considering how to increase the extinction current without loss elsewhere, we are led to the idea of separating the function of triggered breakdown from that of extinction, so as to permit of manipulating the two independently. A tetrode tube is therefore implied, though for reasons other than those obtaining in the case of the stabilizer tube. The corresponding rate meter circuit is shown in fig. 13b.

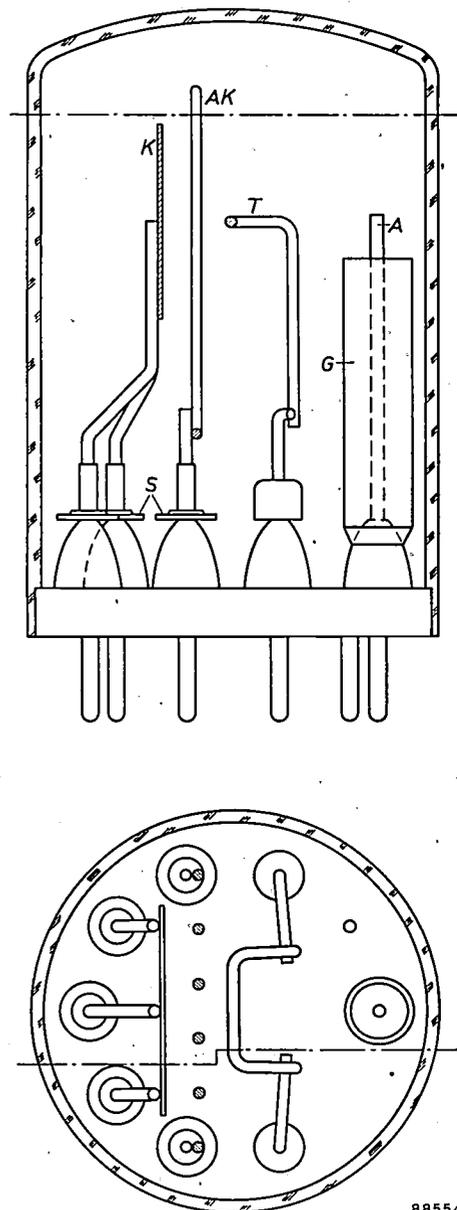


Fig. 14. Simplified diagram of electrode structure of rate meter trigger tube Z 801U. K cathode, AK auxiliary cathode, T trigger electrode, A anode, G glass tube, SS shields.

The first rate meter tetrode had as its fourth electrode ("auxiliary cathode") a perforated disc situated in the path of the discharge to the main cathode (fig. 14). An input signal was now followed by breakdown between anode and auxiliary cathode; then by a moderate discharge leading to a rise in auxiliary cathode potential; then by anode to main cathode breakdown; and finally by the normal downward swing of anode potential. At this point it was found, as expected, that the auxiliary cathode barrier successfully interfered with the transition to a steady glow-discharge between anode and main cathode and led to the collapse of the discharge at higher levels of extinction current than before. Trigger action remained identical with that of a triode, with the auxiliary cathode acting as cathode.

The contribution of the auxiliary cathode to extinction is that of widening the gap between the normal maintaining potential for a steady discharge and the minimum potential reached in the anode "overswing": both as a physical obstacle and as a parasitic current drain, the auxiliary cathode has a greater effect on the former potential.

The tetrode Z 801U has an extinction current of 240  $\mu\text{A}$  against 70 for the triode; the input sensitivity and frequency response are unchanged. In the production version the perforated disc was changed to a wire grille. The tube has a filling of neon with 5% argon, and is shown in fig. 14.

#### The negative pulse circuit

The cold-cathode triode is most frequently switched by a positive signal applied to the trigger; so also were the first tetrode Z 801U tubes. The Geiger-Müller tube however is normally better used with its cathode grounded, and thus delivers a negative pulse. It was found that the tetrode could be triggered by a negative pulse<sup>11)</sup> applied to the auxiliary cathode (fig. 15a), involving a different

sequence of breakdown: the input charge flows across the trigger auxiliary cathode gap; breakdown then occurs between trigger and main cathode; finally, breakdown occurs between anode and main cathode. Apart from its convenience in Geiger-Müller tube service, this negative pulse circuit has three substantial advantages. Firstly there is a remarkable gain in sensitivity: an input charge  $Q_{in}$  of  $3 \times 10^{-11}$  coulomb will suffice, as against the  $5 \times 10^{-10}$  coulomb needed with a positive pulse. Secondly, the input charge  $Q_{in}$  is constant over a wide range of  $V_{as}$ : this makes the sensitivity comparison in practice even more favourable to the negative-pulse circuit. Thirdly, since the function of the auxiliary cathode in initiating breakdown now ends with the passage of the signal charge, there is no objection to inserting an impedance in the auxiliary cathode lead which will prevent instability even if the priming current is equal to  $I_{tb}$ .

The better sensitivity of the negative-pulse circuit follows from reducing the function allotted to the input signal. The essential difference between the two modes of operation is that with the negative pulse applied to the auxiliary cathode, the trigger discharge has to be extended only over the small auxiliary-to-main cathode gap instead of over the much larger anode-to-trigger gap.

This account of the action is supported by the absence of a variation of  $Q_{in}$  with  $V_{as}$ , unlike the positive pulse circuit. Breakdown between trigger and main cathode should be largely unaffected by  $V_{as}$ ; while once this stage has been passed the trigger excursion (about 40 volts) will release  $4 \times 10^{-9}$  coulomb, sufficient to cause anode breakdown for quite small anode supply voltages  $V_{as}$ .

#### Frequency response

The relation between  $R_a$  and anode-circuit time constant  $R_a C_a$  for values of auxiliary-cathode resistor between 4 and 20 M $\Omega$  is such that, for the normal anode swing of 70 volts, an adequate extinction current is obtained even at  $R_a C_a = 300 \mu\text{sec}$ . Against this, the trigger-circuit time-constant is 10 m sec for the values of fig. 15a. This makes the trigger circuit the limiting factor; the value of 10 m sec is too long for a practical radiation monitor. This difficulty was overcome by connecting  $C_t$  to the anode instead of to ground (fig. 15b). If  $C_a \gg C_t$ , then the anode waveform is little affected, while the trigger — after a depression — is restored through positive ion collection by the end of the anode swing to a potential normal for this stage of the cycle. Thereafter anode and trigger potentials rise together and the whole circuit is re-set.

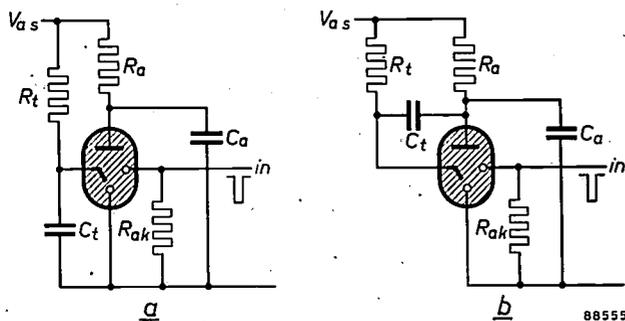


Fig. 15. a) Negative pulse rate meter circuit. b) Negative pulse circuit with improved frequency response. Typical component values:  $R_t = 100 \text{ M}\Omega$ ,  $C_t = 100 \text{ pF}$ ,  $R_a = 200 \text{ k}\Omega$ ,  $C_a = 0.01 \mu\text{F}$ ,  $R_{ak} = 1 \text{ M}\Omega$ ,  $C_{ak} = 15 \text{ pF}$ .

<sup>11)</sup> F.S. Goulding and C.H. Tosswill, British Patent No. 723892.

Until the above circuit change was made it had been assumed that a return to the normal input sensitivity was dependent only upon completion of circuit recovery together with the re-establishment of the priming current. Moreover, it seemed likely that incomplete de-ionization would make the circuit exceptionally responsive for a period. These expectations, however, were not borne out by the results. Fig. 16 shows the variation of the required  $Q_{in}$  with time elapsed since the last input pulse. Circuit equilibrium is restored after about 1 msec, but even after 10 msec sensitivity is still rising slightly.

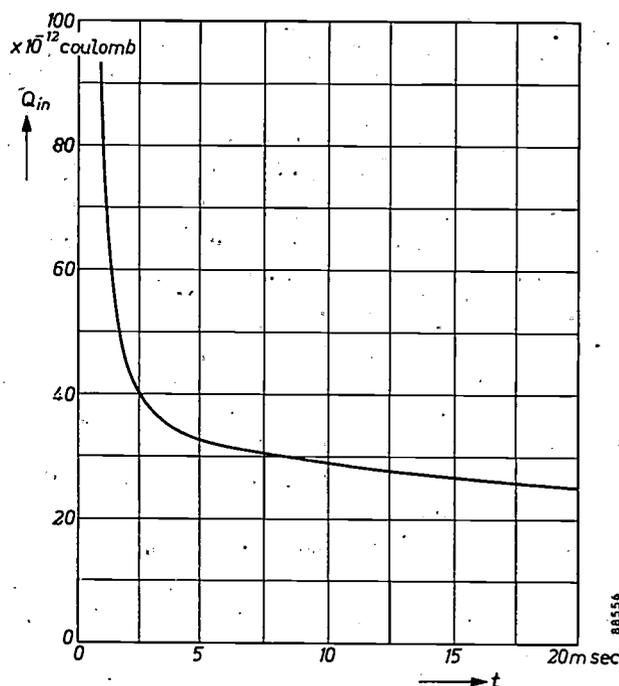


Fig. 16. Recovery of input charge sensitivity. About 1 millisecond after the main discharge has passed, the charge sensitivity of the tube has reached a value of about  $5 \times 10^{-11}$ . At 10 millisecond the sensitivity is  $3 \times 10^{-11}$  and still improving slightly.

The conclusion seems to be that the correct action of the Z 801U tube depends on the development of a very small unstable discharge between trigger and main cathode, and that this unstable discharge cannot occur until the general level of ionization in the tube has fallen too far for the trigger-resistor current  $I_{tb}$  to be provided by ion collection alone. De-ionization in this sense is likely to be protracted, being delayed by the persistence of metastable atoms and by a measure of ion replacement once circuit recovery is complete.

The cure for the delay is evidently to accelerate the process of de-ionization. This has been done in laboratory models having an electrode "labyrinth" in which the discharge passes through narrow passages rather than through a void. However,

as the radiation monitor for which the Z 801U was developed can accept the frequency response provided by the standard structure (up to 1000 c/s), it has proved to be unnecessary to increase the frequency response up to 10 kc/s.

#### Charge amplification

In the rate meter circuit for which the Z 801U was developed, the function of the tube is to deliver a large fixed output charge every time the tube is switched by the arrival of a small input charge not less than  $3 \times 10^{-11}$  coulomb. So far as we are aware, this represents an input sensitivity far higher than that obtainable with any other type of cold-cathode tube. The tube provides, at low frequencies, an amplification factor of not less than  $10^4$ , since  $Q_{out}$  may be as much as  $5 \times 10^{-7}$  coulomb.

If it is desired to make use of input pulses of one volt or less, then the effective input capacitance must exceed 30 pF. A capacitance of this size would upset the stability of the trigger circuit, and must therefore be isolated from the tube by means of selenium diodes<sup>12)</sup>.

The radiation monitor which uses the Z 801U tube as a rate meter tube now also employs another in a variant of the Goulding stabilizer circuit to provide an output of about 155 volts.

The two tubes described above were developed specifically for their intended applications, viz. the stabilizer and rate meter circuits of a radiation monitor. A third tube, the Z803U, has, however, also been developed for more general application. In this tube the priming electrode is an anode, a priming discharge of some 10  $\mu$ A flowing between it and the cathode in a region shielded from the trigger-cathode space. The tube has the same stable triggering characteristics as its predecessors; the nominal trigger breakdown voltage  $V_{tb}$  is 132 V. Anode breakdown voltage is 290 V, while the maintaining voltage  $V_{am}$  is 105 V. The tube can pass a mean current of 8 mA.

Another general purpose tube, type Z70U, of sub-miniature dimensions is in course of development at Eindhoven. This tube can be used with a higher supply voltage than the Z 803U but its mean current capacity is less. A further article will appear in due course in this Review describing these developments.

Acknowledgement should be made of the contribution to this work of Messrs E. Franklin,

<sup>12)</sup> See the article by F. S. Goulding referred to in <sup>8)</sup>.

K. Kandiah and F. S. Goulding of the Electronics Division of A.E.R.E., Harwell, directed by Dr. D. Taylor.

*Summary.* After a brief introduction dealing with the operation and properties of cold-cathode trigger tubes, work is described on the development of two special trigger tubes for use in a radiation monitor. Two of the measures taken to improve reliability and performance are dealt with: the use of molybdenum sputtering to achieve better stability and the provision of a priming current to eliminate the statistical lag. Both tubes are designed to work in self-quenching circuits. The stabilizer

tube, a tetrode with an auxiliary cathode, is designed for providing a stabilizing supply for Geiger tubes. The circuit conditions for satisfactory operation are described. The advantages of this stabilizer are its voltage range (200—300 V), its current range (5—150  $\mu$ A) and the fact that the stabilized output voltage can readily be adjusted by changing the values of two resistors. The rate meter tube is designed to handle Geiger tube pulses. It is also a tetrode but differs from the stabilizer tube in that its priming electrode, again an auxiliary cathode, lies in the path of the main discharge and provides a "direct" priming discharge. The special characteristic of this tube is its extremely high sensitivity; in the negative pulse circuit, the input charge sensitivity is  $3 \times 10^{-11}$  coulomb. This tube can also be used for voltage stabilization.

### AN AUTOMATIC NOISE FIGURE INDICATOR

621.317.34:621.396.822

The output signal from any amplifier is inevitably accompanied by a certain kind of disturbance in the same frequency range as the signal. Amongst the causes of this kind of disturbance are the thermal motion of the electrons in resistors and irregularities in the transit time of electrons in valves. Such disturbances are given the general name of *noise*, irrespective of their origin; their amplitudes have a random distribution, that is, in accordance with a Gaussian curve.

The extent to which an amplifier produces noise is expressed by the *noise factor F*. For a linear amplifier (having an impedance  $Z_i$  between its input terminals)  $F$  is defined, for a frequency  $f$ , as the ratio of the noise-output power  $\Delta P_n$  in the frequency interval  $f \pm \frac{1}{2} \Delta f$  to that part of  $\Delta P_n$  caused by the thermal agitation noise of  $Z_i$  when  $Z_i$  is at a temperature of 290 °K (17 °C).

The noise factor is a function of frequency. To make it possible none the less to characterize the noise properties of an amplifier by a single quantity, the *mean noise factor  $F_m$*  has been introduced. This is by definition <sup>1)</sup> the ratio of the *total* output-noise power  $P_n$  to that part of  $P_n$  caused by the thermal agitation noise of  $Z_i$ ,  $Z_i$  again being at a temperature of 290 °K. Whenever mention is made of the noise factor in what follows, it is the *mean* noise factor  $F_m$  that is meant.

It is often useful to be able to measure the noise factor of, for example, amplifiers in the relay stations of short-wave radio links, or that of equipment for radio-astronomy <sup>2)</sup>. One method of

measuring the noise factor is based on the fact that the power of the thermal noise produced by a resistor is proportional to the absolute temperature of the resistor. In this method the output noise power is measured (with a thermocouple) with two input resistors: first with a resistor that has a suitable resistance at room temperature, and then with a resistor that has the *same* resistance at another known temperature, e.g. that of liquid nitrogen. From the two temperatures, the resistance, and the two output noise powers, the noise factor can then be readily evaluated.

Another well-known method is based on the principle of passing a known noise current through the input resistor, e.g. the current flowing through a saturated diode <sup>3)</sup>. The r.m.s. value of  $i$ , the fluctuations in the current, is determined by

$$\bar{i}^2 = 2eI\Delta f, \dots \dots \dots (1)$$

$e$  representing the charge on the electron and  $I$  the mean value of the diode current. If we connect a resistor  $R_i$ , whose resistance is small compared to the internal resistance of the noise diode, between the input terminals of the amplifier, and then pass through it a noise current of such intensity that the noise power at the output is exactly doubled, the noise figure is given by:

$$F_m = \frac{e}{2kT} IR_i,$$

where  $k$  is Boltzmann's constant and  $T$  the absolute temperature of the resistor  $R_i$ . Inserting the numerical values of the constants  $e$  and  $k$  ( $e = 1.60 \times 10^{-19}$  coulomb,  $k = 1.38 \times 10^{-23}$  joules/°K) and 290 °K

<sup>1)</sup> This definition is in accordance with American Standard 53 IRE 7 S1. Other definitions may be found in the literature; see for example Philips tech. Rev. 2, 329, 1937; 6, 133, 1941; 11, 84, 1949/50.

<sup>2)</sup> See for example C. A. Muller, Philips tech. Rev. 17; 305-315 and 351-361, 1955/56.

<sup>3)</sup> See for example Philips tech. Rev. 2, 330, 1937.

K. Kandiah and F. S. Goulding of the Electronics Division of A.E.R.E., Harwell, directed by Dr. D. Taylor.

*Summary.* After a brief introduction dealing with the operation and properties of cold-cathode trigger tubes, work is described on the development of two special trigger tubes for use in a radiation monitor. Two of the measures taken to improve reliability and performance are dealt with: the use of molybdenum sputtering to achieve better stability and the provision of a priming current to eliminate the statistical lag. Both tubes are designed to work in self-quenching circuits. The stabilizer

tube, a tetrode with an auxiliary cathode, is designed for providing a stabilizing supply for Geiger tubes. The circuit conditions for satisfactory operation are described. The advantages of this stabilizer are its voltage range (200—300 V), its current range (5—150  $\mu$ A) and the fact that the stabilized output voltage can readily be adjusted by changing the values of two resistors. The rate meter tube is designed to handle Geiger tube pulses. It is also a tetrode but differs from the stabilizer tube in that its priming electrode, again an auxiliary cathode, lies in the path of the main discharge and provides a "direct" priming discharge. The special characteristic of this tube is its extremely high sensitivity; in the negative pulse circuit, the input charge sensitivity is  $3 \times 10^{-11}$  coulomb. This tube can also be used for voltage stabilization.

## AN AUTOMATIC NOISE FIGURE INDICATOR

621.317.34:621.396.822

The output signal from any amplifier is inevitably accompanied by a certain kind of disturbance in the same frequency range as the signal. Amongst the causes of this kind of disturbance are the thermal motion of the electrons in resistors and irregularities in the transit time of electrons in valves. Such disturbances are given the general name of *noise*, irrespective of their origin; their amplitudes have a random distribution, that is, in accordance with a Gaussian curve.

The extent to which an amplifier produces noise is expressed by the *noise factor*  $F$ . For a linear amplifier (having an impedance  $Z_i$  between its input terminals)  $F$  is defined, for a frequency  $f$ , as the ratio of the noise-output power  $\Delta P_n$  in the frequency interval  $f \pm \frac{1}{2} \Delta f$  to that part of  $\Delta P_n$  caused by the thermal agitation noise of  $Z_i$  when  $Z_i$  is at a temperature of 290 °K (17 °C).

The noise factor is a function of frequency. To make it possible none the less to characterize the noise properties of an amplifier by a single quantity, the *mean noise factor*  $F_m$  has been introduced. This is by definition <sup>1)</sup> the ratio of the *total* output-noise power  $P_n$  to that part of  $P_n$  caused by the thermal agitation noise of  $Z_i$ ,  $Z_i$  again being at a temperature of 290 °K. Whenever mention is made of the noise factor in what follows, it is the *mean* noise factor  $F_m$  that is meant.

It is often useful to be able to measure the noise factor of, for example, amplifiers in the relay stations of short-wave radio links, or that of equipment for radio-astronomy <sup>2)</sup>. One method of

measuring the noise factor is based on the fact that the power of the thermal noise produced by a resistor is proportional to the absolute temperature of the resistor. In this method the output noise power is measured (with a thermocouple) with two input resistors: first with a resistor that has a suitable resistance at room temperature, and then with a resistor that has the *same* resistance at another known temperature, e.g. that of liquid nitrogen. From the two temperatures, the resistance, and the two output noise powers, the noise factor can then be readily evaluated.

Another well-known method is based on the principle of passing a known noise current through the input resistor, e.g. the current flowing through a saturated diode <sup>3)</sup>. The r.m.s. value of  $i$ , the fluctuations in the current, is determined by

$$\bar{i}^2 = 2eI\Delta f, \dots \dots \dots (1)$$

$e$  representing the charge on the electron and  $I$  the mean value of the diode current. If we connect a resistor  $R_i$ , whose resistance is small compared to the internal resistance of the noise diode, between the input terminals of the amplifier, and then pass through it a noise current of such intensity that the noise power at the output is exactly doubled, the noise figure is given by:

$$F_m = \frac{e}{2kT} IR_i,$$

where  $k$  is Boltzmann's constant and  $T$  the absolute temperature of the resistor  $R_i$ . Inserting the numerical values of the constants  $e$  and  $k$  ( $e = 1.60 \times 10^{-19}$  coulomb,  $k = 1.38 \times 10^{-23}$  joules/°K) and 290 °K

<sup>1)</sup> This definition is in accordance with American Standard 53 IRE 7 S1. Other definitions may be found in the literature; see for example Philips tech. Rev. 2, 329, 1937; 6, 133, 1941; 11, 84, 1949/50.

<sup>2)</sup> See for example C. A. Muller, Philips tech. Rev. 17; 305-315 and 351-361, 1955/56.

<sup>3)</sup> See for example Philips tech. Rev. 2, 330, 1937.

for  $T$ , we find:

$$F_m = 20.0 I R_i \dots \dots \dots (2)$$

(For frequencies above about 100 Mc/s a correction to formulae (1) and (2) is necessary in view of the transit time effect; absolute calibration with a resistor as noise source is then advisable.)

Instead of doubling the output noise power by the addition of diode noise, another possibility is the temporary reduction of the voltage amplification of the amplifier under test by a factor of  $\sqrt{2}$ , whereupon the output noise power is brought back to its initial value by the addition of diode noise. The power of the added noise is again a measure of the noise factor (provided of course that the variation of the amplification is effected at a point of the circuit where it cannot perceptibly influence the noise factor or the bandwidth).

This last method might be carried out as follows: the bias of the output valve of the amplifier (the output valve, because it contributes hardly anything to the output noise) is made to alternate periodically between the value at which the amplification is  $A$  and the value at which the amplification is  $A/\sqrt{2}$ ; we shall call these intervals the odd and the even half-periods respectively. Simultaneously with this variation of the amplification, the anode voltage of the diode is modulated in such a way that there is no diode current in the odd half-periods and the saturation current  $I$  in the even half-periods. The diode current can be controlled by altering its filament current, for it possesses a directly heated tungsten cathode. The filament current is now so adjusted that the demodulated output voltage of the amplifier becomes as low as possible. Once this condition is reached, the added diode noise just compensates the  $\sqrt{2}$ -fold decrease in the amplification; in accordance with (2) the diode current  $I$  is now a measure of the noise factor sought.

A direct-reading noise factor meter could be built on this principle, but its operation would be a fairly complicated business, since in each case the filament current would have to be adjusted by manual control to a value such that the demodulated output voltage was at a minimum. If the task is to bring down an amplifier's noise figure to the minimum, every alteration to a network parameter would necessitate measuring the noise factor in the manner described above, and if three or more network parameters had to be altered, the job of finding the optimum setting might be a matter of several days. The whole procedure could be considerably speeded up if there was an instrument available that would automatically indicate the noise factor without requiring any

manual adjustment, thus giving a direct indication of the effect of every alteration in the amplifier<sup>4</sup>).

This aim may be achieved by providing the above-mentioned arrangement with *automatic* regulation of the filament current to the value at which the demodulated output voltage is at a minimum<sup>5</sup>). Fig. 1 shows curves for this voltage according as the diode current  $I$  is (a) too low, (b) at the right value and (c) too high; it is a square-wave voltage of 400 c/s (this is the frequency selected for varying the amplification and for modulation of the diode), which passes through zero and changes its phase when  $I$  passes the correct value. In order to obtain an effective system of control, the filament current for the noise-producing diode is likewise given a frequency of 400 c/s, and supplied by an auxiliary amplifier controlled by the demodulated output voltage. (For this a *selective* amplifier tuned to 400 c/s is necessary, since only the component of 400 c/s must be amplified.) If  $I$  is too small, then the filament current, and hence also  $I$ , are automatically increased; if  $I$  is too large, the filament current is reduced.

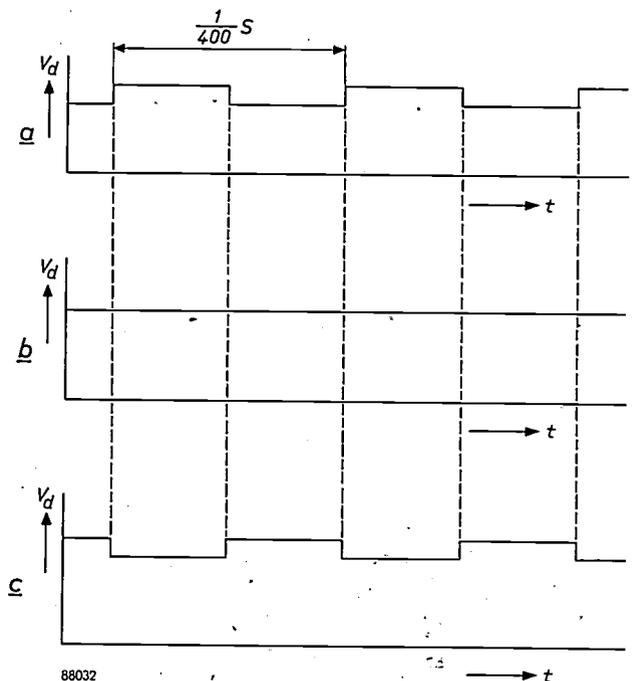


Fig. 1. Demodulated output voltage when the filament current of the noise-producing diode is (a) too low, (b) of the right value, and (c) too high. The frequency with which the amplification is varied and the noise diode is modulated is 400 c/s.

<sup>4</sup>) Our thanks are due to Mr. C. Seeger (Netherlands Foundation for Radio Astronomy, Leiden) for drawing our attention to the importance of such an instrument and for discussing its initial design with us.

<sup>5</sup>) Since the conclusion of this work two papers have been published (American patent 2 691 098 of October 5, 1954, in the name of W. Selove, and an article by H. Wallmann in Acta Polytechnica (El. Eng. Series) 6, 1955, No. 6), in which instruments for the same purpose are described, with minor differences of design.

As in most automatic control systems, the "actual value of the variable quantity" is compared with the "desired value", whilst the difference between the two serves for adjusting the "control element". In our case the actual value of the initial quantity is the output voltage during the even half-periods; the desired value is the output voltage during the odd half-periods; the difference is the block-shaped demodulated output voltage; and the control element is the noise-producing diode.

In principle any value may be selected for  $R_i$ , provided that it is small compared to the internal resistance of the noise-producing diode.  $50 \Omega$  is suitable, since it is a standardized value in all equipment for decimetric and centimetric waves in most parts of the world. With  $R_i = 50 \Omega$ , equation (2) shows that the noise figure is just equal to twice the diode current in mA read on the meter (this current flows only during half the time, so that  $I$  is twice the meter reading).

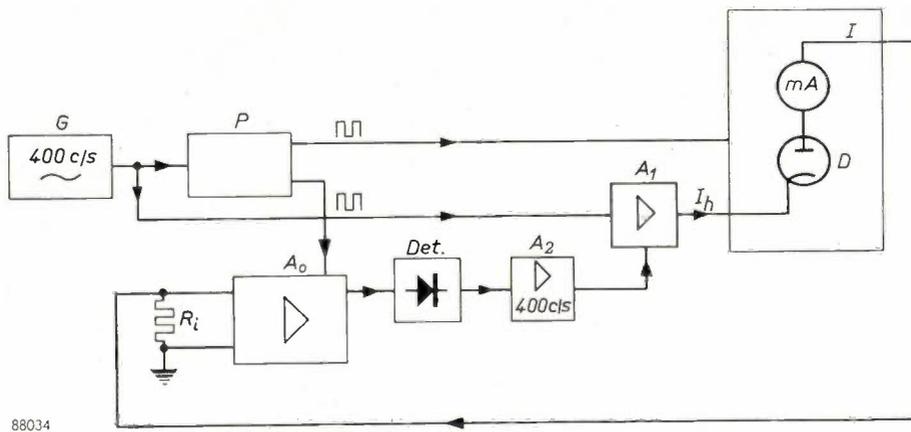


Fig. 2. Block diagram of the automatic noise factor indicator. The RC generator  $G$  produces a sinusoidal voltage of 400 c/s.  $A_1$  amplifier producing the filament current  $I_h$  for the noise diode  $D$ .  $P$  pulse shaper supplying a square-wave voltage of 400 c/s.  $A_0$  amplifier whose mean noise factor is to be measured.  $R_i$  input resistance.  $Det$  detector.  $A_2$  selective amplifier, which passes only a narrow band at 400 c/s and controls the amplification of  $A_1$ . The mean noise factor is read on the milliammeter  $mA$ .

The working principle is shown in the block diagram of fig. 2.  $G$  is an RC generator, supplying a sinusoidal voltage of 400 c/s. The amplifier  $A_1$  connected to it supplies the filament current  $I_h$  for the noise-producing diode  $D$ . Also connected to  $G$  is a pulse shaper  $P$ , in which the sinusoidal voltage is amplified and clipped, producing a square-wave voltage. The main function of this voltage is to give the periodic variation in the amplification of  $A_0$ , the amplifier under test, between the values  $A$  and  $A/\sqrt{2}$ ; this is effected by applying the square-wave variable-amplitude voltage to the cathode of the last tube of  $A_0$ . The square-wave voltage also effects the synchronized opening and closing of the noise-producing diode. During the even half-periods the latter produces a noise current through the input resistance  $R_i$  of the amplifier  $A_0$ . The selective amplifier  $A_2$  amplifies only the 400 c/s component of the output voltage demodulated by the detector  $Det$ , and regulates the amplifier  $A_1$  in such a way that it produces approximately the correct filament current. The correct value is approached the more closely as the overall amplification in the closed loop  $A_0$ - $Det$ - $A_2$ - $A_1$ - $D$ - $A_0$  is greater.

Fig. 3 is a photograph of an instrument constructed on the principles described above.



Fig. 3. Automatic noise factor indicator constructed in accordance with the block diagram of fig. 2. (The instrument is not being manufactured.)

As regards the accuracy of the measurements given, this is in principle limited by the fact that the output voltage during the even half-periods can only be made to approximate to that during the odd half periods, since both are noise voltages. The error thus arising can be reduced by narrowing the bandwidth  $\beta$  of the selective amplifier  $A_2$  (fig. 2). (Strictly speaking, it should be taken into account that the time constant of the filament of the diode is a contributory factor in determining the bandwidth after detection.) Thus, when a noise spectrum, assumed to be of rectangular shape (bandwidth  $\Delta$ ), is demodulated and the result is passed through a filter (bandwidth  $\beta \ll \Delta$ ), we obtain a voltage whose r.m.s. value  $V$  shows the relative fluctuation <sup>6)</sup>

$$\frac{\Delta V}{V} = \frac{\pi^{\frac{3}{2}}}{2\sqrt{2}} \sqrt{\frac{\beta}{\Delta}} \approx 2 \sqrt{\frac{\beta}{\Delta}}$$

In amplifiers for decimetric waves (for which our instrument has been primarily developed)  $\Delta$  may be, say,  $5 \times 10^6$  c/s; if  $\beta = 5$  c/s,  $\sqrt{\beta/\Delta}$  is of the

<sup>6)</sup> This formula follows from calculations by R. H. Dicke, *Rev. sci. Instr.* 17, 268-275, 1946.

order of  $10^{-3}$ , so that the error does not exceed a few tenths percent.

For frequencies above 1000 Mc/s, the high-vacuum diode with tungsten filament is not the most suitable noise source. In such cases it is better to use a gas-discharge tube <sup>7)</sup>, e.g. of the types K 50 A or K 51 A, which are specially developed for noise measurements. With these tubes the noise cannot be controlled by the filament current. The circuit described above, however, may be modified so that, instead of the filament current, the ratio of the duration of the odd to that of the even intervals is controlled. The even intervals (during which noise is added) should then be made so long that they correspond to a noise factor greater than that to be measured. Since the gas-discharge tube has to operate continuously, one must employ either an amplifying tube with an alternately amplifying and suppressing action (gating tube), or an attenuator with electrically varied attenuation. Further details concerning these two possibilities will not be dealt with here.

<sup>7)</sup> K. S. Knol, *Philips Res. Rep.* 6, 289-302, 1951.

F. L. H. M. STUMPERS and N. van HURCK.

# Philips Technical Review

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

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

## FERROXPLANA \*), HEXAGONAL FERROMAGNETIC IRON-OXIDE COMPOUNDS FOR VERY HIGH FREQUENCIES

by G. H. JONKER, H. P. J. WIJN and P. B. BRAUN.

621.318.134

*Among high-frequency soft magnetic materials, large-scale use is made of ferroxcube. The upper frequency limit of ferroxcube lies at about 100 Mc/s. New groups of soft magnetic materials have now been developed in the Philips Laboratories at Eindhoven, which may be used at considerably higher frequencies. Like ferroxcube and ferroxdure, these materials are prepared by the sintering of metal oxides. In this article the composition and some of the properties of the new materials are discussed.*

The behaviour of ferroxcube materials at high frequencies is briefly as follows<sup>1)</sup>. For a number of applications the *initial permeability*  $\mu_i$ , i.e. the permeability in a weak magnetic field, is important. For a given type of ferroxcube,  $\mu_i$  remains constant up to a certain limiting frequency, which may lie somewhere between 1 Mc/s and 100 Mc/s. If the frequency is raised above this limit,  $\mu_i$  decreases and the magnetic losses show a sharp increase.

The product of the initial permeability at low frequency and the frequency at which this permeability begins to decrease is practically the same for the various types of ferroxcube. Snoek<sup>2)3)</sup> has explained this as follows. Considered magnetically, the material is divided into a large number of small regions, viz. the Weiss domains. Each domain is magnetized to saturation and this magnetization is bound by a certain restoring couple to a given preferred direction. If, due to some cause, the direction of magnetization is temporarily deflected from the preferred direction, then, owing to the restoring couple and to the fact that the magnetic moment is associated with a mechanical angular momentum (electron spin), the magnetization

vector will precess about the preferred direction with a frequency  $f_r$  proportional to the strength of restoring couple<sup>4)</sup>. If an external alternating magnetic field is applied, a resonance effect occurs as the frequency approaches  $f_r$ , which is responsible for the above-mentioned large magnetic losses and the decreased permeability. On the other hand there is the fact that in ferroxcube materials the initial permeability at low frequency  $\mu_i$  — being presumably mainly due to a rotation of the magnetization vector in each Weiss domain — is likewise essentially determined by the strength of the restoring-couple binding the magnetization to the preferred direction. The susceptibility  $\chi_i$  at low frequency, which is equal to  $\mu_i - 1$ <sup>†)</sup>, is inversely proportional to this restoring couple, so that the product  $\chi_i f_r$  may be expected to be a constant, virtually independent of the type of ferroxcube. The value of this product as derived from a quantitative study is in good agreement with the value found experimentally.

\*) Ferroxlana, with the *a* pronounced as in the word plane.

1) See e.g. J. J. Went and E. W. Gorter, Philips tech. Rev. **13**, 181-208, 1951/52.

2) J. L. Snoek, Nature **160**, 90, 1947.

3) J. L. Snoek, Physica **14**, 207-217, 1948.

4) For a more detailed explanation, see H. G. Bèljers and J. L. Snoek, Philips tech. Rev. **11**, 313-322, 1949/50.

†) *Editorial note:* This is valid in the rationalized Giorgi system of units. In the c.g.s. electromagnetic system  $\chi_i = (\mu_i - 1)/4\pi$ . In this article  $\mu_i$  is used to represent the *relative* initial permeability with respect to that of free space  $\mu_0$ . Quantities expressed in the c.g.s. system will be placed between square brackets. In the rationalized Giorgi system the magnetization  $M$  is given by the formula:

$$B = \mu_0 (H + M).$$

The relation between  $\chi_i$  and  $f_r$  imposes an absolute frequency limit on the application of the various ferroxcubes as a soft magnetic material. The frequency at which the material is to be used must lie below  $f_r$ . Because of the constant value of the product  $\chi_i f_r$ , very high values of  $f_r$  imply that the initial permeability will be only of the order of unity, so that there is no point in using this material.

In this article we shall deal with new ferromagnetic materials which, like ferroxcube, are non-metallic in nature and of high resistivity and can therefore be used at high frequencies. In these materials, also, the magnetization has certain preferred directions. In some of these materials, however, the strength of the restoring couple depends also upon the *direction* in which the magnetization is deflected from its position of equilibrium. For a similar case Snoek<sup>3)</sup> pointed out the theoretical possibility of evading the above relation between  $\chi_i$  and  $f_r$ . This is indeed possible with the new materials; among them, in fact, are soft magnetic materials that can be used up to far higher frequencies than ferroxcube; they possess e.g. a frequency-independent initial permeability of approximately 10 up to about 1000 Mc/s. It is proposed to coin the word *ferroxplana* as a generic name for these materials. The significance of this name will reveal itself in a later stage of this article.

Let us first examine the chemical composition of the new compounds. We shall then deal with their magnetic properties, with particular attention to the phenomenon of ferromagnetic resonance in the specific case where the magnetization in different directions is bound to varying degrees with the position of equilibrium.

#### Chemical composition of the new compounds

The new materials are compounds of the oxides BaO, Fe<sub>2</sub>O<sub>3</sub> and MeO, Me representing one of the small divalent ions of Mn, Fe, Co, Ni, Zn or Mg, or a mixture of these. The materials are prepared by heating specific mixtures of these oxides for a given time at a temperature of 1200-1400 °C, so that a sintered ceramic product is obtained.

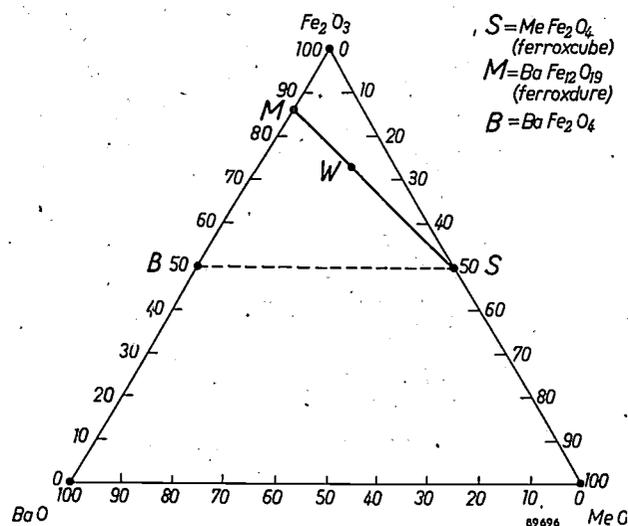
The composition of these compounds can be represented in a triangle, the vertices of which are formed by the constituent oxides (*fig. 1*). The points S, M, and B on the sides of this triangle represent the following known compounds:

S: MeFe<sub>2</sub>O<sub>4</sub>, the ferrites with the cubic spinel structure (ferroxcube).

M: BaFe<sub>12</sub>O<sub>19</sub>, with the hexagonal magnetoplumbite structure, the main component of the permanent magnet material ferroxdure.

B: BaFe<sub>2</sub>O<sub>4</sub>, a non-ferromagnetic compound.

In earlier work on permanent magnet materials, the compound BaFe<sup>II</sup><sub>2</sub>Fe<sup>III</sup><sub>16</sub>O<sub>27</sub> was found within the above-mentioned triangle<sup>5)</sup>. This compound is represented in *fig. 1* by the point W on the line connecting M and S, Me in the latter representing an



*Fig. 1.* The composition of the W compounds is represented by a point within the triangle. The symbol Me represents one of the divalent ions Mn, Fe, Co, Ni, Zn or Mg or a mixture of these. The known materials ferroxcube (S), ferroxdure (M) and BaFe<sub>2</sub>O<sub>4</sub> (B) are represented by points on the sides of the triangle.

Fe<sup>II</sup>-ion. The presence of this compound led to investigations to see whether a series of corresponding compounds could be prepared in which other Me-ions were substituted for the Fe<sup>II</sup>-ions. (A similar substitution is possible with Fe<sup>II</sup>Fe<sup>III</sup><sub>2</sub>O<sub>4</sub>, resulting in S-type compounds, the ferroxcubes.)

The formation of such "substituted" W compounds can be studied by means of X-ray powder diffraction studies of the final product. Since, however, M and W are structurally closely related (see<sup>5)</sup>), small quantities of M which can appear as a second phase due to incomplete chemical reaction cannot be easily identified in the X-ray diagrams (see *fig. 2*). This difficulty can be avoided as explained below.

Like the M compounds, the W compounds are ferromagnetic with a hexagonal crystal structure. The preferred direction of magnetization is parallel to the hexagonal axis. The small crystals in powdered materials will therefore assume such an orientation in a magnetic field that the hexagonal axis of each crystal becomes more or less parallel to the lines of force. A few drops of a suspension of

<sup>5)</sup> H. P. J. Wijn, *Nature* **170**, 707, 1952.  
P. B. Braun, *Nature* **170**, 708, 1952.

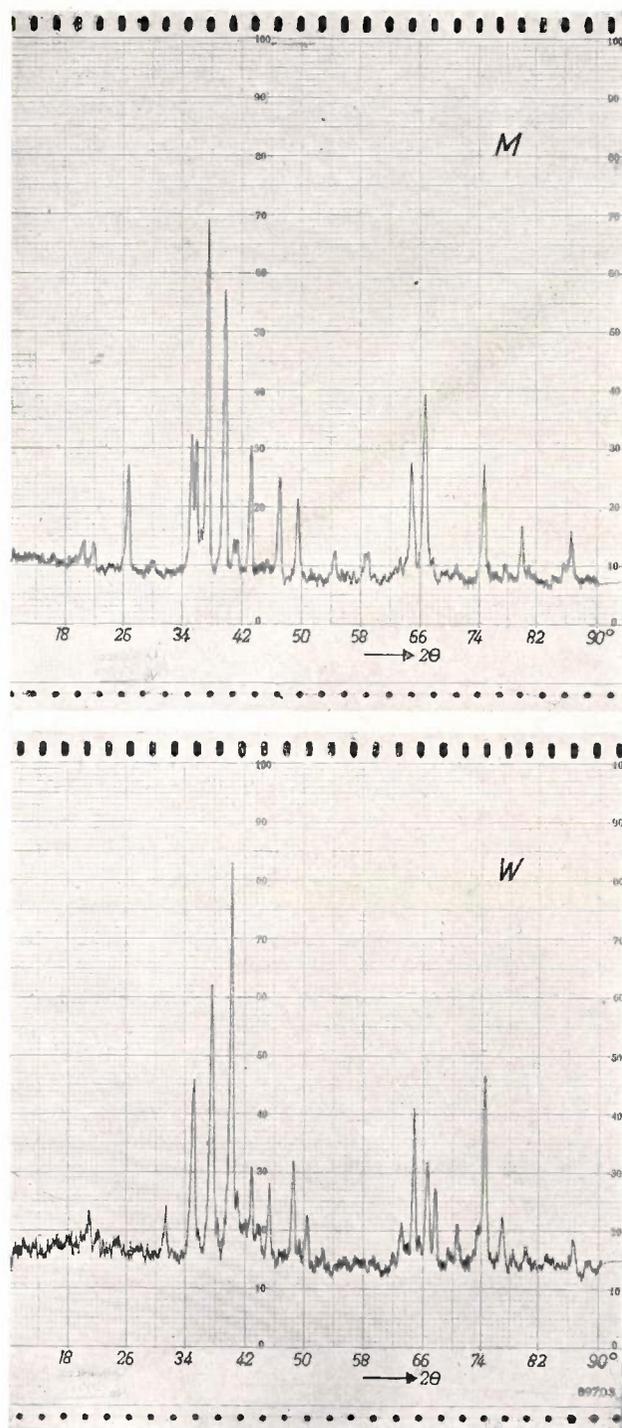


Fig. 2. X-ray diffraction diagrams of powders of M and W compounds, taken with Co K $\alpha$  radiation.

the powdered specimen in a solution of a binder are now applied on a glass plate, after which a strong magnetic field is applied perpendicular to this plate. When the solvent of the binder is allowed to evaporate completely before the magnetic field is switched off, the crystals will remain with their hexagonal axes perpendicular to the glass surface. In the X-ray diffraction pattern of such

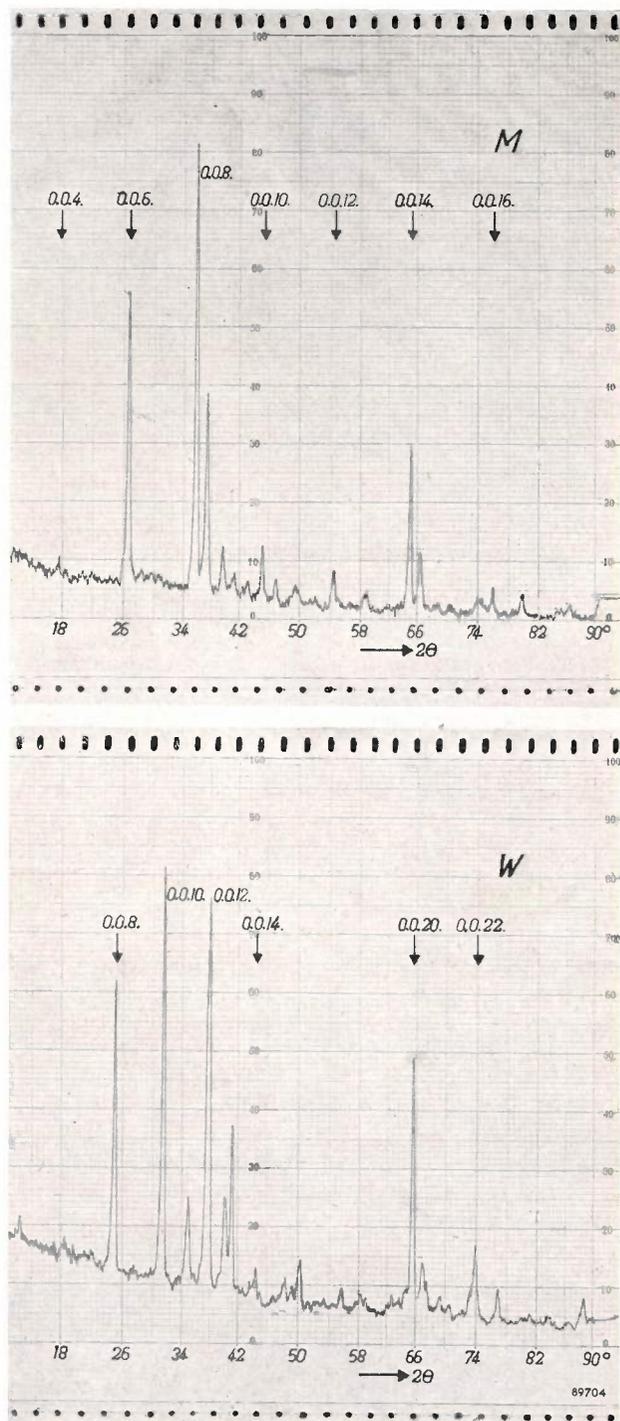


Fig. 3. X-ray diffraction diagram of magnetically oriented powder specimens of M and W. The reflections from the basal planes 0.0.l are intensified with respect to those of fig. 2.

a specimen obtained with the diffractometer<sup>6)</sup>, the X-ray reflections from the planes perpendicular to the hexagonal axis appear much intensified, whereas the other reflections have nearly disappeared. This greatly simplifies the diagram, see fig. 3.

<sup>6)</sup> The diffractometer is well-suited to the investigation of flat specimens such as this; cf. W. Parrish, E. A. Hamacher and K. Lowitzsch, Philips tech. Rev. 16, 123-133, 1954/55.

These simplified diagrams are characteristically different for M and W, so that now a small quantity of one of these compounds can be easily identified in the presence of a large quantity of the other compound.

This method makes it possible to examine with great accuracy the reaction products formed by the sintering of the oxide mixtures. With this sensitive measure of the reaction, it was possible to improve the process such that pure compounds could be obtained, in which the  $\text{Fe}^{\text{II}}$ -ions are entirely or partly replaced by other ions. We shall henceforth designate the separate compounds of this group, which is generally indicated by the symbol W, by prefixing the W with the symbols of the divalent ions in question. NiZnW, for instance, stands for the compound  $\text{BaNiZnFe}_{10}\text{O}_{27}$ .

A general examination of the reaction products obtained from mixtures of  $\text{BaO}$ ,  $\text{Fe}_2\text{O}_3$  and  $\text{MeO}$  revealed that other ferromagnetic compounds are also formed. The X-ray diagrams of these compounds were found to be so similar to those of M and W that, with the aid of them alone, it was impossible to establish with what compositions of oxide mixtures they could be precisely synthesized. The method of orientation in a magnetic field was therefore again used, and the result was once more a characteristic simplification of the diffraction patterns. This made it possible to establish that *two* more compounds occur, indicated in the triangle by the points Y and Z (see fig. 4), with the chemical formulae:



and

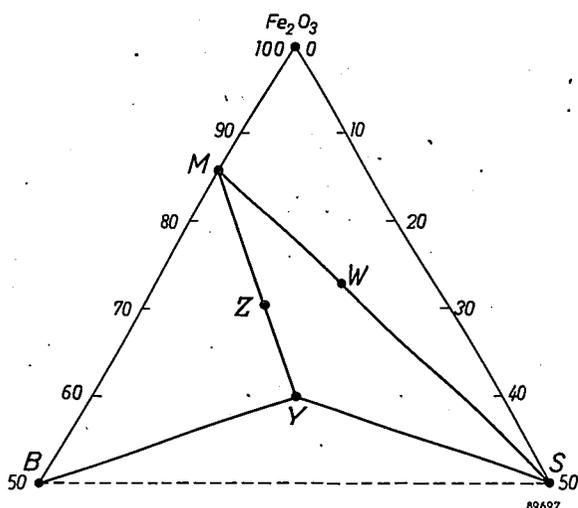


Fig. 4. Situation of the points for the new compounds Y and Z in the top section of the triangle of fig. 1.

Further investigations showed that these compounds possess hexagonal crystal structures closely related to those of M and W (see Appendix).

If the various divalent ions are substituted for Me, we obtain compound groups, which will be called the Y group and the Z group. The separate compounds in these groups will be designated in a manner similar to those in the W group.

#### Preferred direction and preferred plane

The diffraction patterns of the magnetically oriented specimens (fig. 5) show that the crystals of the Z and the Y specimens have a different orientation relative to the magnetic field. For most Z specimens the hexagonal axis (*c*-axis) is found to be parallel to the magnetic field; the spontaneous magnetization of these compounds is therefore directed along the *c*-axis. This axis is a *preferred direction*, just as with the M and the W group. With all Y specimens on the other hand the hexagonal axis is perpendicular to the magnetic lines of force, the directions of the other axes being yet arbitrary; this means that the spontaneous magnetization of the Y compounds lies in a plane *perpendicular* to the *c*-axis. In this plane, which we call the *preferred plane*, the magnetization can have virtually any direction.

Further investigation has revealed that a preferred plane does not occur solely with the Y group, but also with those compounds of the Z and W groups in which Co-ions are completely or partly substituted for Me. Table I indicates for all new compounds in which one specific type of ions are substituted for Me, whether a preferred direction ( $\uparrow$ ) or a preferred plane ( $\perp$ ) occurs. The occurrence of a preferred plane is of great consequence as regards the magnetic behaviour, as we shall see later. For the various groups of iron oxide compounds with a preferred plane of magnetization the generic name of ferroplana is proposed. Table II gives the saturation magnetization and the Curie temperature of some of the new materials, both with a preferred plane and with a preferred direction.

The occurrence of a preferred plane can also be expressed more quantitatively. If a preferred direction occurs in a hexagonal structure, then the energy *E* required per unit volume to deflect the

<sup>7)</sup> This might be written in simpler form by halving the number of atoms, thus obtaining  $\text{BaMeFe}_6\text{O}_{11}$ . Nevertheless the above formula is preferred, owing to the fact that, like those of the other compounds discussed, it indicates the correct number of atoms per "structural unit", as will be seen later in this article.

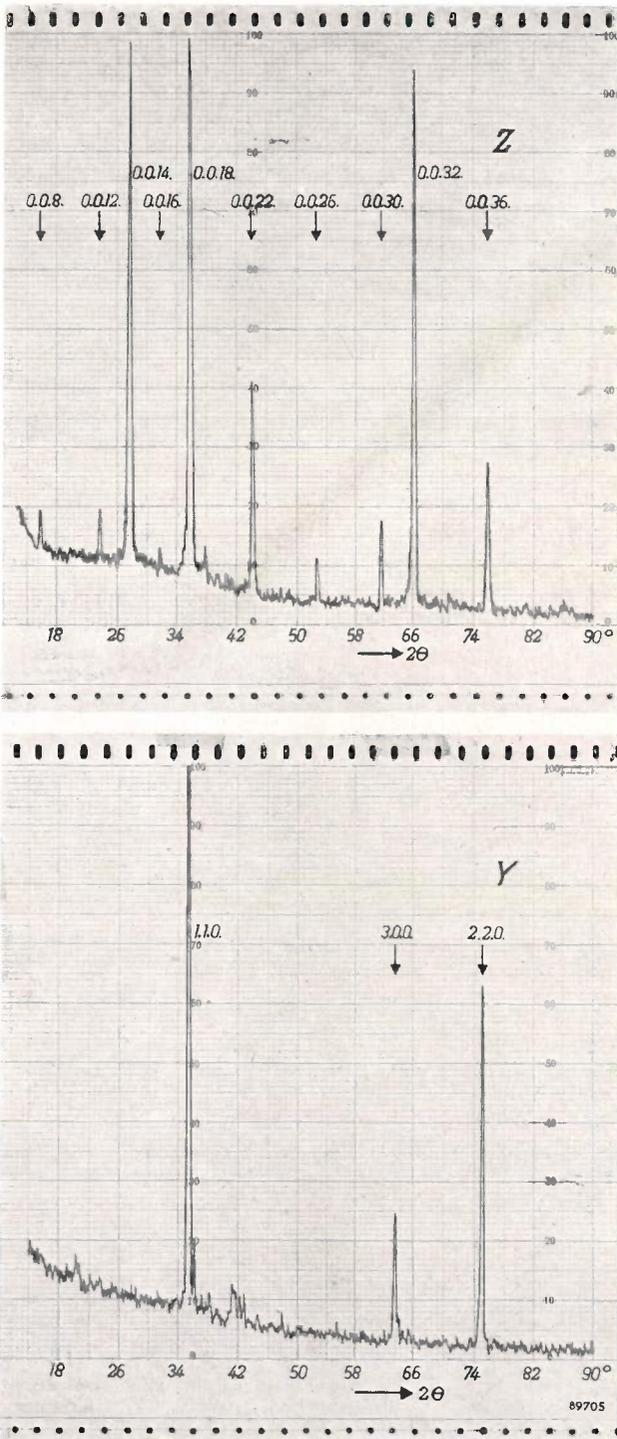


Fig. 5. X-ray diffraction diagrams of magnetically oriented powder specimens of Z and Y. With the Z compounds the reflections from the 0.0.l planes (basal planes) are again intensified; with Y compounds, reflections from the h.k.0 planes (parallel to hexagonal axis) are intensified.

magnetization through an angle  $\theta$  with respect to the direction of the hexagonal axis, can be roughly expressed by <sup>8)</sup>:

$$E = K \sin^2\theta. \quad (1)$$

<sup>8)</sup> See J. J. Went, G. W. Rathenau, E. W. Gorter and G. W. van Oosterhout, Philips tech. Rev. 13, 194-208, 1951/52.

Table I. Compounds with a preferred direction ( $\uparrow$ ) or a preferred plane ( $\perp$ ).

Group	Divalent ion substituted for Me					
	Mn	Fe	Co	Ni	Zn	Mg
W	$\uparrow$	$\uparrow$	$\perp$	$\uparrow$	$\uparrow$	$\uparrow$
Z	$\uparrow$	$\uparrow$	$\perp$	$\uparrow$	$\uparrow$	$\uparrow$
Y	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$

The quantity  $K$  is termed the crystal anisotropy constant. We see that for positive  $K$ , the direction  $\theta = 0$ , i.e. the  $c$ -axis, is the direction of minimum energy. In the case of a preferred plane the energy minimum occurs for  $\theta = 90^\circ$ , which means that  $K$  in (1) is now negative. We can thus express the occurrence of a preferred plane by saying that the compounds in question have a negative crystal anisotropy. We could also say that in such a case the direction  $\theta = 0$  is an "abhorred direction" for the magnetization.

Table II. Saturation magnetization  $\mu_0 M$  at 20 °C and Curie temperature  $T_c$  for some of the new materials. For comparison the values of  $\mu_0 M$  and  $T_c$  for ferrocube 4E are also given.

Material	$\mu_0 M$ in $10^{-4}$ Wb/m <sup>2</sup> [4 $\pi M$ in gauss]	$T_c$ in °C
Mn <sub>2</sub> W	4000	410
CoFeW	4500	430
ZnFeW	4800	422
Ni <sub>0.5</sub> Zn <sub>0.5</sub> Fe <sub>1</sub> W	4600	425
Co <sub>2</sub> Y	2900	340
Mg <sub>2</sub> Y	1400	300
Co <sub>2</sub> Z	3500	400
Ferrocube 4E	3280	595

If we replace Co in the W or Z group by a mixture of Co and an increasing quantity of another ion, e.g. Zn, then we notice that at a certain percentage of Zn the preferred plane changes into a preferred direction. The crystal anisotropy, initially negative, gradually changes with increasing Zn-content towards a positive value.

One may also consider the crystal anisotropy energy of the cubic ferromagnetic materials, such as ferrocube (spinel structure). Because of the cubic structure, however, the energy  $E$  must now be invariant with respect to the edges of the cube, the latter being equivalent. Let  $a_1, a_2$  and  $a_3$  be the direction cosines of a given direction with respect to the three crystal axes, then the simplest formula to replace (1) is;

$$E = K' (a_1^2 a_2^2 + a_2^2 a_3^2 + a_3^2 a_1^2) \quad (2)$$

This formula is an approximation. We can derive from it that  $K'$  is positive when the edges of the cube are the preferred directions and negative when the body diagonals are the preferred directions.

With these ferrites we likewise find that the Co-substitution gives rise to a change in the sign of  $K'$ :  $K'$  is positive for cobalt ferrite<sup>9)</sup> and negative for all other ferrites. If we compare the structures of the cubic and the hexagonal compounds and consider both as the close-packing of spheres of oxygen ions, then the direction of one of the body diagonals of the cubes is equivalent to the  $c$ -axis of the hexagonal structure. If these are the preferential directions for each structure (and this is true for both types of compounds if no Co is present, apart from the Y-group compounds), then the anisotropy is positive in the hexagonal substances, but negative in the cubic ones. The difference in sign for the anisotropy when comparing cubic and hexagonal compounds arises from the definition of  $K$  and  $K'$  in (1) and (2) respectively. Since these definitions have become generally adopted, we shall continue to use them, in spite of the possible confusion which may arise.

### The importance of materials with preferred plane

The fields of application of these hexagonal compounds depend on whether they have a preferred direction or a preferred plane.

For compounds with a *preferred direction*, the corresponding crystal anisotropy is as a rule substantially greater than that of cubic compounds (ferroxcube). The magnetization is then rigidly bound to that direction, which renders such materials in principle suitable for application as a permanent magnet material, with properties similar to ferroxdure (M).

For compounds with a *preferred plane*, the magnetization is virtually free to assume any direction in that plane, or at least the direction is not determined by the value of  $K$ , which only specifies the energy required for a rotation *out of* the plane. There are, of course, other forms of crystal anisotropy, and moreover the stress anisotropy and the shape anisotropy (see<sup>1)</sup>), which cause a preferred direction of magnetization within the plane and oppose rotation within the plane, but these are as a rule small (they also occur in ferroxcube). Ferroplana materials are therefore unsuitable for use as permanent magnets. On the other hand, they may in principle have a large permeability and are thus suitable as soft magnetic materials. Owing to their non-metallic character, these materials have fields of application similar to the ferroxcubes, particularly for high frequencies. It is found that the ferromagnetic resonance effect, which places a frequency limit on the application

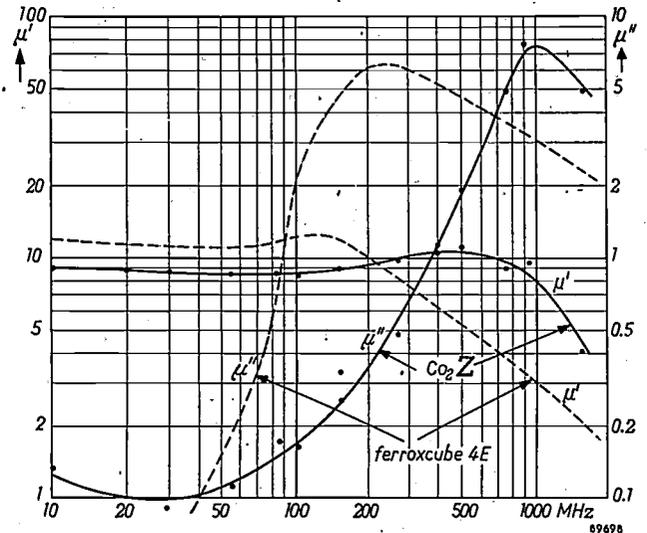


Fig. 6. The real part  $\mu'$  and the imaginary part  $\mu''$  of the initial permeability  $\mu_i$ , plotted as a function of the frequency  $f$  for the materials  $\text{Co}_2\text{Z}$  (full lines) and the cubic nickel ferrite ferroxcube 4E (dotted lines).

of magnetic materials, occurs with many ferroplana materials only at far higher frequencies than with ferroxcube. As an example, *fig. 6* shows the experimentally determined initial permeability  $\mu_i$  of the compound  $\text{Co}_2\text{Z}$  plotted as a function of the frequency. The permeability of the cubic nickel ferrite, ferroxcube 4E is also shown. The real part  $\mu'$  and the imaginary part  $\mu''$  of the initial permeability (see e.g.<sup>1)</sup>) are given;  $\mu''$  is proportional to the losses, and the loss angle  $\delta$  is given by  $\tan \delta = \mu''/\mu'$ . The diagram shows that, although the permeability at low frequency is about the same for both materials, the value of  $\mu'$  for  $\text{Co}_2\text{Z}$  at 1000 Mc/s hardly differs from that at low frequency, whereas the value of  $\mu'$  for ferroxcube 4E at 1000

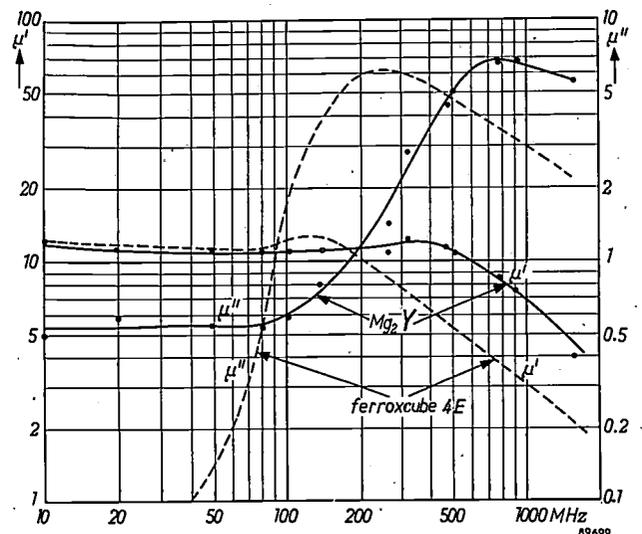


Fig. 7.  $\mu'$  and  $\mu''$  as functions of  $f$  for the materials  $\text{Mg}_2\text{Y}$  (full lines) and ferroxcube 4E (dotted lines).

<sup>9)</sup> R. M. Bozorth, E. F. Tilden and A. T. Williams, *Phys. Rev.* **99**, 1788-1789, 1955.

Mc/s has nearly dropped to unity. The difference in ferromagnetic resonance frequency can also be seen in the curves for  $\mu''$ . If, for example, the frequencies are compared at which  $\mu''/\mu' = 0.1$ , it can be seen that  $\text{Co}_2\text{Z}$  can be used up to a frequency considerably higher (e.g.  $5\times$ ) than ferroxcube 4E.

Fig. 7 similarly compares ferroxcube 4E with the compound  $\text{Mg}_2\text{Y}$ . Here too, we find the same characteristic of the initial permeability of  $\text{Mg}_2\text{Y}$  remaining constant to a higher frequency. Other ferroxplana materials show similar results.

In the following section this behaviour of ferroxplana will be considered theoretically.

**Ferromagnetic resonance in materials with a preferred plane**

*The resonance frequency*

In the introduction it has been shown that the ferromagnetic resonance in cubic ferrites depends on the tightness with which the magnetization is bound to the preferred direction and to the fact that a magnetic moment is fundamentally associated with a mechanical angular momentum. This can be quantitatively analysed as follows. The couple tending to restore the magnetization to the preferred direction can be described for small deviations from this direction by introducing an apparent internal magnetic field  $H_A$  in this direction, the anisotropy field, which is given by:

$$H_A = \frac{2K}{\mu_0 M}$$

$M$  representing the saturation magnetization. That this field is equivalent to the anisotropy can be seen from the fact that the couple  $\vec{\mu}_0 M \times \vec{H}_A$  occurring when the magnetization is turned through a small-angle  $\theta$  from the preferred direction, has the appropriate value of  $2K\theta$ . This couple is equal to the rate of change in the angular momentum, so that

$$\vec{M} \times \vec{H} = \frac{\dot{\vec{M}}}{\gamma} \quad (3)$$

where  $\gamma$  is the gyromagnetic ratio, i.e. the ratio of the magnetic moment to the corresponding angular momentum (0.221 Mc/s per A/m. [17.6 Mc/s per Oe]). From this the resonance frequency  $f_r'$  can be easily derived:

$$2\pi f_r' = \gamma H_A \quad (4)$$

Let us now consider the more complex resonance in materials with a preferred plane. The situation is illustrated by fig. 8, in which the  $xy$ -plane represents the preferred plane. Let us assume that the magnetization within this plane has an additional (weak) preference for the  $x$ -direction, which may be described as an anisotropy field  $H_{A1}$ . This anisotropy refers to the case where  $M$  is turned out of the  $x$ -direction towards the  $y$ -direction, i.e. within the preferred plane, as shown in fig. 8a. If  $M$  is turned towards the  $z$ -direction, so that it is moved out of the preferred plane (fig. 8b), the far greater anisotropy field  $H_{A2}$  is involved.

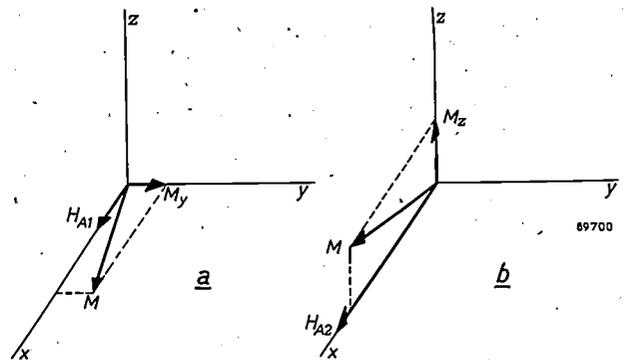


Fig. 8. a) If the magnetization is deflected within the preferred plane ( $xy$ -plane) the anisotropy field  $H_{A1}$  is involved. b) A deflection out of the preferred plane involves the far stronger field  $H_{A2}$ .

Under the influence of the two anisotropy fields a deviation of  $M$  from the  $x$ -direction will again cause it to describe a precession with a certain as yet unknown frequency  $f_r$ . The situation of fig. 8a can be represented by:

$$\gamma M_y H_{A1} = \dot{M}_z \quad (5)$$

whilst the situation of fig. 8b can be put as:

$$\gamma M_z H_{A2} = -\dot{M}_y \quad (6)$$

Differentiating (5) and substituting for  $\dot{M}_y$  from (6), we obtain

$$\gamma^2 H_{A1} H_{A2} M_z + \ddot{M}_z = 0 \quad (7)$$

whilst a similar equation can be derived for  $M_y$ . The frequency is now given by:

$$2\pi f_r = \gamma \sqrt{H_{A1} H_{A2}} \quad (8)$$

$f_r$  is thus proportional to the geometrical mean of the anisotropy fields<sup>3) 10)</sup>. If these fields are equal, we clearly arrive at formula (4) again.

<sup>10)</sup> C. Kittel, Phys. Rev. 73, 155-161, 1948.

It is apparent from (8) that the resonance frequency for materials with a preferred plane can be far higher than that for materials having a cubic structure. The anisotropy field  $H_{A1}$  for deviation of the magnetization *within* the preferred plane is of the same order of magnitude as the  $H_A$  of ferroxcubes, whereas the anisotropy field for a deflection *out of* the preferred plane is generally much larger than  $H_{A1}$ .

In figs. 6 and 7 it was shown that ferroxlana not only possesses a far higher resonance frequency than ferroxcube, but also has about the same initial permeability. A theoretical study, briefly reproduced in the following section, leads to the same conclusion.

*The initial permeability*

In the general case that the coupling of the magnetization to a preferred direction can be described in terms of two anisotropy fields  $H_{A1}$  and  $H_{A2}$ , the susceptibility of a crystal in the direction of the three main axes (see fig. 8) can be written as:

$$\chi_{xx} = 0; \chi_{yy} = \frac{M}{H_{A1}}; \chi_{zz} = \frac{M}{H_{A2}}$$

These expressions are based on the assumption that for these materials, as for ferroxcube, the susceptibility is primarily governed by rotation processes. For a polycrystalline material with a random orientation of the crystallites, and disregarding their magnetic interaction, the resultant susceptibility is given by:

$$\chi = \frac{1}{3} \left( \frac{M}{H_{A1}} + \frac{M}{H_{A2}} \right) \dots \dots (9)$$

Materials such as ferroxcube and also ferroxdure form a special case where  $H_{A1} = H_{A2} (= H_A)$ , so that formula (9) is simplified into

$$\chi_{\uparrow} = \frac{2}{3} \frac{M}{H_A} \dots \dots (10a)$$

The large initial permeability of ferroxcubes is attributable to the fact that their  $H_A$  is small (about  $10^2 - 10^4$  A/m [ $1 - 100$  Oe]).

For the materials with a preferred plane described here,  $H_{A2} \gg H_{A1}$ ,  $H_{A1}$  being comparable in magnitude to  $H_A$  of ferroxcube, whilst  $H_{A2}$  is of the same order of magnitude as  $H_A$  of ferroxdure (approx.  $10^6$  A/m [approx.  $10^4$  Oe]). A good approximation is thus obtained by:

$$\chi_{\perp} = \frac{1}{3} \frac{M}{H_{A1}} \dots \dots (10b)$$

This formula shows that also with ferroxlana substantial permeability values may occur in principle, since  $M$  assumes values comparable to those of ferroxcube (see table II).

Finally we shall illustrate the essential difference in behaviour between ferroxlana and ferroxcube in a manner similar to that used by Snoek (for single crystals)<sup>3</sup>). Combining (8) and (9) produces the formula:

$$2\pi f_r \chi = \frac{2}{3} \gamma M \left( \frac{1}{2} \sqrt{\frac{H_{A1}}{H_{A2}}} + \frac{1}{2} \sqrt{\frac{H_{A2}}{H_{A1}}} \right) \dots (11)$$

The factor between brackets always amounts to unity for ferroxcubes; this implies that for a given value of the permeability approximately the same limiting frequency is found for all ferroxcubes, as pointed out in the introduction. For the case of  $H_{A2} \neq H_{A1}$ , such as with ferroxlana, the factor between brackets is always *greater* than unity, to a degree that is dependent upon  $H_{A2}/H_{A1}$ . Whenever the permeability values are about equal, this factor immediately indicates to what extent the frequency limit of ferroxlana is raised with respect to that of ferroxcube.

*Comparison between theory and experiment*

Formula (8) could be tested experimentally for various materials because they could be obtained in the form of single crystals, in which  $H_{A2}$  could be directly measured;  $H_{A1}$  can be evaluated by means of eq. (10b) from the quantities  $\chi_{\perp}$  and  $M$ , measured on the sintered materials.

Table III gives a comparison of the values of the resonance frequency obtained from (8) with those measured experimentally. The experimental resonance frequency was defined as the frequency at which the permeability has dropped to half the value at low frequency. The agreement between the theoretical and the measured values may be considered satisfactory.

Table III. Theoretical and measured values of the resonance frequency for two materials.

Material	$\chi_{\perp}$ *) at low freq.	$\mu_0 M$ *) in $10^{-4}$ Wb/m <sup>2</sup> [ $4\pi M$ in gauss]	$H_{A1}$ in A/m	$H_{A2}$ in A/m	$f_r$ in $10^3$ Mc/s	
					computed from (8)	measured
Co <sub>2</sub> Z	7	2700	$10^4$	$1.1 \times 10^6$	3.7	2.5
Mg <sub>2</sub> Y	9	900	$0.25 \times 10^4$	$1.2 \times 10^6$	2.1	1.0

\*) The values for  $\chi_{\perp}$  and  $M$  refer to porous sintered materials.

### Appendix: The crystal structure of the compounds of the W, Y and Z groups

The crystal structure of the new compounds was determined by means of an X-ray investigation. We shall not enter into the investigation itself<sup>11)</sup>, but the results can be briefly summarized as follows. Like M and W, the Y and Z compounds have a hexagonal crystal structure, which can be described as the close-packing of spheres of the large oxygen and barium ions. Layers of these ions are perpendicular to the hexagonal axis. In the unit cell each layer contains four large ions. In most layers these are four oxygen ions, but at regular intervals there is a layer with three oxygen ions and one barium ion. In M compounds these barium-containing layers occur once in five layers, and in W compounds once in seven layers. In Y compounds two adjacent layers each contain a barium ion; these double layers are then followed by four oxygen

“structural unit”. From these “structural units” or “blocks” of five, seven and six layers respectively, which we have termed the  $M_5$ ,  $W_7$  and  $Y_6$  blocks, the unit cells of M, W, Y and also Z can be built up as follows.

The unit cells of M, W, and Y can be built up by stacking together two  $M_5$ , two  $W_7$  and three  $Y_6$  blocks respectively. In doing this, the blocks must be so turned that the crosses indicated in fig. 9 become centres of symmetry. By placing the block  $Y_6$  in the orientation as it is shown here upon the block  $M_5$  we obtain a block that we shall call  $Z_{11}$ . Two  $Z_{11}$  blocks then form the unit cell of Z, provided they are joined in such a way that the centres of symmetry in one of the boundary layers coincide. Finally it may be observed that the  $W_7$  block can be considered as a stacking of an  $M_5$  block with two layers of the S composition. The essentials of the data given above are summarized in Table IV.

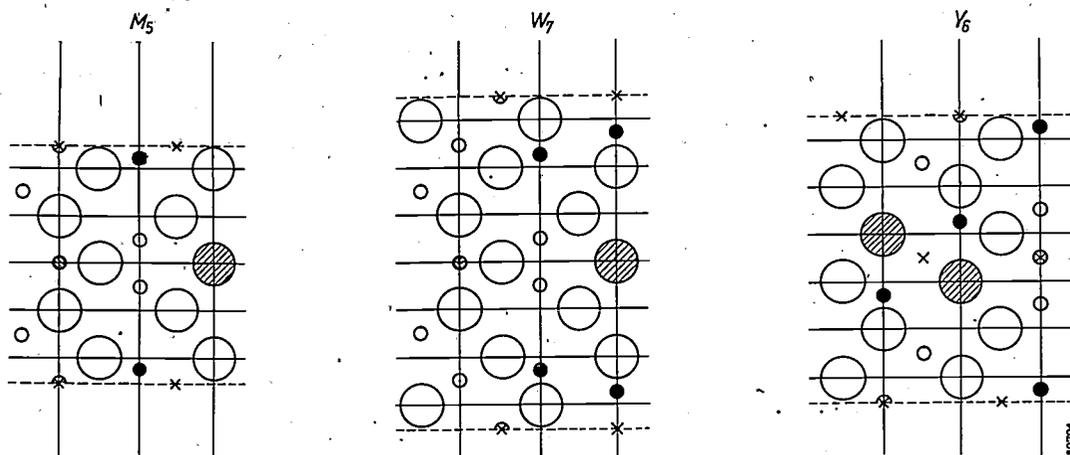


Fig. 9. Cross-section through a plane of symmetry of the “structural units” or “blocks”  $M_5$ ,  $W_7$  and  $Y_6$ . The crosses represent centres of symmetry, the vertical lines are threefold axes of symmetry. All atoms not situated on any of these axes thus occur three times in each block. The complete spatial structure of a block can be obtained by rotating this cross-section e.g. about the central axis of symmetry  $120^\circ$  forward and  $120^\circ$  backward out of the plane. We thus get the layers of four large ions of oxygen (○) or barium (●) as mentioned in the text. These layers are indicated by horizontal lines. The small black dots represent metal ions (Me) at the tetrahedral sites, the small white circles are metal ions at the octahedral sites, whilst the small hatched circles represent the metal ions surrounded by five oxygen ions.

The diagram shows that the chemical formulae for M, W and Y given above indeed represent the number of ions contained in  $M_5$ ,  $W_7$  and  $Y_6$  blocks.

layers. The small metal ions are found in the interstices in the oxygen lattice. Just as with the spinel structure, these are either tetrahedral sites, surrounded by four oxygen ions, or octahedral sites, surrounded by six oxygen ions. In M and W compounds an additional metal ion occurs in the Ba-containing layer, at a site surrounded by five oxygen ions. These structures can be visualized with the aid of fig. 9. This shows a cross section through a plane of symmetry of the structures. In the  $c$ -direction the cross-section does not extend over the entire length of the unit cell, but over part of it, a so-called

Table IV. Chemical formulae and some crystallographical properties of the hexagonal ferromagnetic compounds dealt with.

Group	Formula	Unit cell		
		$a$ -axis in Å	$c$ -axis in Å	Structure
M	$BaFe_{12}O_{19}$	5.9	23.3	$2M_5$
W	$BaMc_2Fe_{16}O_{27}$	5.9	32.8	$2W_7 = 2(M_5 + 2S)$
Y	$Ba_2Mc_2Fe_{12}O_{22}$	5.9	43.5	$3Y_6$
Z	$Ba_3Mc_2Fe_{24}O_{41}$	5.9	56.0	$2(M_5 + Y_6)$

<sup>11)</sup> P. B. Braun, Crystal structures of a new group of ferromagnetic materials, Philips Res. Rep., to be published shortly.

The values measured for the saturation magnetization of the new compounds are in good agreement with the crystal structures described above. Néel<sup>12)</sup> demonstrated that in ferromagnetic oxides with spinel structure (the S-compounds) the magnetic moment of the various ions can be divided into two groups. The magnetic moments in one group have parallel orientation, but are anti-parallel to the magnetic moments of the other group (cf. 1)). The total magnetic moment is then equal to the difference between the resultant magnetic moments of both groups. This so-called ferrimagnetism has also been found in the compound M and likewise occurs in the compounds of the W, Y and Z groups. Just as with the S-compounds, the total magnetic moment depends upon the ions substituted for Me. Measurements have shown that the magnetic moment per gram-molecule of a given Z-compound is equal to the sum of the moments per gram-molecule of M and the corresponding Y compounds. The magnetization can thus be evaluated with the aid of the formula:

$$Z_{11} = M_5 + Y_6,$$

which describes the composition of the  $Z_{11}$  blocks. It was further found that the magnetization per gram-molecule of a given W compound is in general equal to that of M added to the magnetization of two gram-molecules of the corresponding S-material, so that

$$W_7 = M_5 + 2S,$$

<sup>12)</sup> L. Néel, Ann. Physique 3, 137-198, 1948.

which is in agreement with the structure of the  $W_7$  block given above.

It is fortunate that the moments of the blocks composing the unit cells are all parallel and not anti-parallel. Owing to this, the compounds W, Y and Z have saturation magnetizations comparable to those of the ferroxcubes.

**Summary.** A number of new ferromagnetic materials with interesting properties are described. These materials are compounds of BaO, Fe<sub>2</sub>O<sub>3</sub> and MeO, Me representing a divalent ion from the series Mn, Fe, Co, Ni, Zn, Mg. They are prepared by a ceramic process and are related, as regards crystal structure, to ferroxdure. The chemical formulae and the crystal structures are then discussed. The crystal structures are hexagonal and can be described as the close-packing of spheres of oxygen ions, ranged in layers perpendicular to the c-axis, in which at certain sites a barium ion is substituted for an oxygen ion. Small metal ions are situated in the interstices between the oxygen ions. A number of "structural units" or "blocks" constitute the elements from which the unit cells of all the new compounds can be built up.

Like ferroxdure, some of these materials show a preferred direction for the magnetization, parallel to the c-axis. In other compounds, for which the generic name ferroxplana has been coined, this direction is an "abhorred" direction; here the magnetization has a preferred plane perpendicular to the c-axis. Inside this plane the magnetization direction is relatively arbitrary, so that these materials may possess a permeability of approximately 10, about the same as that found—below 100 Mc/s—for the cubic nickel ferrite, ferroxcube 4E. As a consequence of the presence of a preferred plane, the ferromagnetic resonance frequency of ferroxplana lies far higher than that of ferroxcube. These new materials can therefore be used as soft magnetic materials up to a frequency about 5× higher than ferroxcube 4E.

## A TRANSISTOR RADIO RECEIVER POWERED BY A THERMOPILE

621.396.62:621.375.4:621.362

A problem that has long exercised radio designers is the construction of a receiver capable of reproducing music at normal room volume without requiring any power supply from the mains or from batteries. Even in highly developed areas of the world such as Europe and the U.S.A. many isolated farms and cottages have no electric supply mains, and in less developed areas a mains supply is the exception rather than the rule. Dry batteries as a source of supply are generally rather expensive and have the additional disadvantage of having a limited storage life, particularly in tropical regions. Very few people can afford the luxury of a private generating plant with petrol or diesel engine.

It is not therefore surprising that the *thermopile* has been widely considered as a possible source of supply. The thermopile consists of a series of wires or bars of two dissimilar conductors A and B welded end to end alternately. All the junctions A-B are heated to a temperature  $T_1$  and all junctions B-A are kept at a temperature  $T_2$ . The thermoelectromotive force then developed is roughly proportional to the temperature difference  $T_1 - T_2$ . The e.m.f. per °C temperature difference and per junction is termed the thermoelectric power, a quantity depending on the materials A and B. There is no simple rule for calculating the thermoelectric power, but in general we find a thermoelectric power of some  $\mu\text{V}/^\circ\text{C}$  for combinations of pure metals, up to some tens of  $\mu\text{V}/^\circ\text{C}$  for certain alloys and up to some hundreds of  $\mu\text{V}/^\circ\text{C}$  for some semiconductors.

In the past, the power efficiency of the thermopile was the main consideration, and its usefulness was judged accordingly. This efficiency, however, which is not higher than about 7% at best, may sometimes play only a subordinate part in deciding the usefulness of a thermopile, as is the case here.

Experiments were made some years ago in the Philips Research Laboratories in Eindhoven with thermopiles consisting of a number of thermocouples of easily weldable and non-corrosive alloys (e.g. "Chromel"- "Alumel"). These piles, when heated by a gas flame, produced an e.m.f. of about 2 V, whilst the internal resistance was about 2  $\Omega$ . This was sufficient to heat the filaments of the sub-miniature valves with which a radio receiver had been equipped; the anode voltage was obtained by means of a vibrator converting the voltage of 2 V into one of about 100 V. For a consumption of 1 W it was

thus possible to attain a volume of sound amply sufficient for an ordinary living-room.

The advent of the transistor has simplified matters considerably. It has no filament and requires a supply of only a few volts, so that a vibrator is no longer necessary. Hence relatively few thermocouples are required for the pile. The pile should preferably be of simple design and unlikely to break down, whilst if it does break down it should be easily repairable by, say, a local radio mechanic or smith. The construction of such a thermopile, made up of a few hundred pieces of wire of not too fine a gauge, then becomes a reasonable proposition.

To recapitulate, the problem may be formulated as follows: is it possible to make a satisfactory combination of a cheap heat source, a thermopile, and a radio receiver with transistors and loud-speaker? Laboratory experiments, which will be outlined below, have shown that such a combination is feasible.

The heat source that comes immediately to mind is the paraffin lamp, since this is widely used as a source of light in houses having no electricity or gas supply. Moreover, it is possible, by means of a thermopile, to utilize that part of the heat of a paraffin lamp that is otherwise simply wasted. The present experiments were carried out with a lamp equipped with a Welsbach incandescent mantle. For a consumption of about  $\frac{1}{8}$  pint of paraffin per hour this lamp produces an agreeable white light of about 125 candle power; according to present prices in the Netherlands, it works out that the light costs about the same per lumen as that from an electric filament lamp. The special shape of the lamp-glass allows the draught to be regulated so that the vapourized paraffin burns quietly under the mantle, without any formation of soot.

The thermopile has to be adapted to the lamp so as to give good heat transfer but it must not disturb the draught. One of the thermopiles at present being used is made up of "rosettes", units of twelve thermocouples in a radial arrangement. The pile as a whole is a stack of sixteen such rosettes placed on top of the (shortened) lamp-glass (see *figure*). All the thermocouples, numbering less than 200 in all, are connected in series by spot-welds. They consist of "Chromel"-constantan, a combination with a thermoelectric power of about 65  $\mu\text{V}$  per °C temperature difference. The wire is 1.2 mm thick; the internal resistance  $R_i$  of the entire pile

is  $5 \Omega$ . Half the junctions point inwards and are in the stream of hot gases ascending from the chimney; the other half point outwards and are cooled by the surrounding air. When the lamp is burning with a normal flame, the average temperature difference between the hot and the cold junctions is about

the terminal voltage is 1.1 V at a current of 220 mA. This terminal voltage is rather low for transistors which have to drive a loudspeaker. The power produced, on the other hand, is higher than necessary; 100 mW is ample. The terminal voltage then becomes 1.9 V, which is a suitable value.



Left: Paraffin lamp (make "Aladdin") fitted with incandescent gas mantle. The thermopile is placed on the shortened lampglass. The metal tube above serves to maintain a good draught. Right: Experimental radio receiver with 7 transistors and one germanium diode, supplied from the thermopile.

$180^\circ\text{C}$  in our model, so that an electromotive force  $E$  of about 2.2 V is produced. The power  $P$  produced across an external resistance  $R$  amounts to

$$P = \frac{R}{R + R_i} \times \frac{E^2}{R + R_i}$$

and is a maximum when  $R = R_i$ . For  $E = 2.2 \text{ V}$  and  $R = R_i = 5 \Omega$  the power is equal to 242 mW. This means that for maximum power consumption

For the sake of simplicity the receiver has been designed as a "straight set", consisting of a single-stage R.F. amplifier with an OC 45 transistor, a detector stage with a germanium diode OA 79 of the point-contact type<sup>1)</sup>, and a five-stage A.F. amplifier<sup>2)</sup>. Four of these five stages comprise an

<sup>1)</sup> Philips tech. Rev. 16, 225-232, 1954/55.

<sup>2)</sup> If the supply voltage were 6 V instead of about 2 V, higher collector resistances could be used, which would improve amplification per stage to such an extent that one stage could be omitted.

OC 71 transistor; the output stage employs two OC.72 transistors in push-pull arrangement. In view of the experimental character of the set, A.G.C. has not been incorporated, nor have special provisions been made to compensate for large variations in ambient temperature.

Of the power of 100 mW to be supplied by the thermopile, obviously only a portion is available as A.F. power for the loudspeaker. This portion, however, is larger than might be expected, viz. about 50%. An A.F. power of 50 mW, applied to a sensitive loudspeaker (efficiency about 8%) produces sufficient sound volume for an ordinary room.

For minimum power consumption, with a given audio output power, the output stage is best arranged as a class B push-pull amplifier (i.e. the transistors are just cut off in the absence of a signal). This, however, would involve considerable distortion of weak signals. For this reason an arrangement intermediate between class B and class A was preferred (transistors not entirely cut off in the absence of a signal).

If, between the two bases (input electrodes) of the push-pull output stage, an audio voltage is applied of such amplitude that the transistors operate at full modulation, the collectors (output electrodes) of this stage together pass an average current of about 50 mA. The remainder of the set consumes about 10 mA, so that the total current will be about 60 mA and the terminal voltage 1.9 V. The same applies for normal operation of the receiver on a radio signal with a modulation depth of 100%. Nearly always, however, the modulation depth of the signal is considerably less, with the result that less current is consumed; on the average this is only 40 mA.

If no special provisions were made, the fairly high internal resistance of the thermopile would give rise to feedback, since this resistance would be common to all stages; the receiver would therefore be liable to oscillation (at a frequency of a few c/s). To prevent this, two measures have to be taken:

- 1) the coupling between the various stages via the common supply source has to be reduced, e.g. by shunting the thermopile with a large capacitance;

- 2) the amplification for very low frequencies has to be reduced by employing a coupling capacitor of low capacitance in one of the stages.

An inductive aerial of the "Ferroceptor" type<sup>3)</sup>, i.e. a coil wound on a ferroxcube rod acting as a frame aerial, is used. In the present case the rod is 200 mm long and 15 mm thick, the coil has a self-inductance  $L = 185 \mu\text{H}$  (a value suitable for medium waves), the effective height<sup>4)</sup>  $h = 12 \text{ mm}$  and the  $Q$ , without load, is 160. The receiver produces its maximum output power with a signal of field strength  $F = 1 \text{ mV/m}$  and having a modulation depth of 30%.

If the aerial is matched to a resistance  $R_a$ , it produces in this resistance a power  $P_a = (QFh)^2/4R_a$ . Writing  $R_a = 2\pi f LQ$ ,  $f$  being the frequency of the incoming radio signal, we find, for  $f = 10^6 \text{ c/s}$ ,  $P_a = 5 \times 10^{-12} \text{ W}$ . To obtain an output power of 50 mW we must effect a total amplification of  $50 \times 10^{-3}/5 \times 10^{-12} = 10^{10}$  times, i.e. 100 dB. Assuming a modulation depth of 30% and in view of the fact that the efficiency of the detector stage is considerably below 100%, the overall amplification should in fact amount to about 125 dB. The receiver described here gives just this amplification.

A common touchstone for the quality of reproduction of a receiver is the R.F. and I.F. bandwidth. As a measure of this we may take the difference between the two frequencies at which the amplification has fallen to 1/1.6 of its value at the resonance frequency.

In this receiver the bandwidth so measured is 4.25 kc/s. The attenuation on detuning by 9 kc/s (i.e. the frequency separation between broadcasting transmitters in the medium wave range) amounts to a factor of 8. This figure is fairly low in comparison with that of ordinary radio receivers, but could only be improved by employing more complicated circuitry. In general, however, it will suffice for interference-free reception of powerful and not too distant transmitters, which was the purpose we had in mind when developing this experimental receiver.

J. van HENGEL and J. VOLGER

<sup>3)</sup> H. Blok and J. J. Rietveld, Philips tech. Rev. 16, 186-188 and 191-193, 1954/55.

<sup>4)</sup> See the article quoted in <sup>3)</sup>, 190 and 193.

## THE APPLICATION OF FERROXCUBE IN UNIDIRECTIONAL WAVEGUIDES AND ITS BEARING ON THE PRINCIPLE OF RECIPROCITY

by H. G. BELJERS.

621.372.852.223 : 621.318.134 : 538.566.5

*Until recently, the application of ferroxcube was limited to high frequency cores or other magnetic circuits, up to frequencies (10 to 100 Mc/s) at which the losses became appreciable. It has been found, however, that even at much higher frequencies (1000 to 30 000 Mc/s), where the magnetic losses may be very great, there are important possibilities of application, especially for magnetically polarized ferrites, such as ferroxdure and magnetized ferroxcube. In this article, the theory and practice of such applications will be examined more closely.*

### The principle of reciprocity

The principle of reciprocity<sup>1)</sup> is manifested in many branches of physics, such as, for example, mechanics, optics and electricity. It may be formulated as follows: if a cause  $X$ , acting on a point  $A$  of a system, brings about an effect  $x$  at a point  $B$ , then the same cause  $X$ , acting on  $B$ , will give rise to the same effect  $x$  at  $A$ .

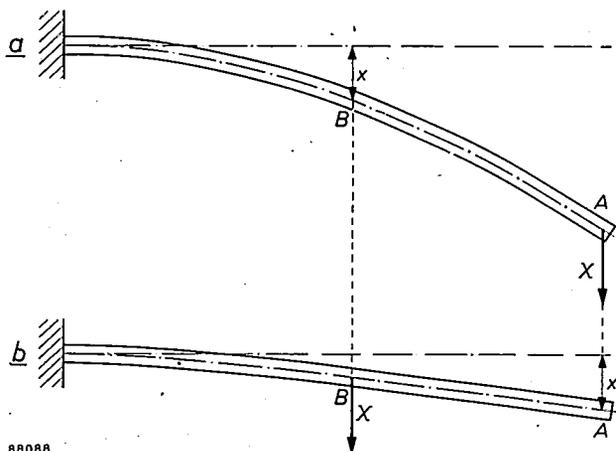


Fig. 1. A horizontal rod clamped at one end, as an example of the principle of reciprocity in elasticity. A force  $X$  at  $A$  causes a displacement  $x$  at  $B$ . The same force  $X$  at  $B$  causes an identical displacement  $x$  at  $A$ .

Fig. 1 illustrates a simple example from elasticity. A rod clamped at one end in a horizontal position, is loaded at its other end with a weight  $X$ , as a result of which point  $B$  is given a vertical displacement  $x$ . If the same weight  $X$  acts at  $B$ , then  $A$  will likewise be displaced through a distance  $x$  (notwithstanding the fact that the shape of the bent rod differs in the two cases).

An electrical example is furnished by a linear network (composed of resistances, capacitances and inductances). If, in such a system, an electromotive force  $X$ , acting in a branch  $A$  of the network causes a current  $x$  in a branch  $B$ , the e.m.f.  $X$  acting in branch  $B$  will give rise to the same current  $x$  in branch  $A$ ; this is true not only for direct voltages and currents, but also for alternating voltages and currents: in the latter case the currents  $x$  in  $A$  and  $B$  will not only be of the same amplitude — they will also have the same phase.

The principle of reciprocity has been used in studying the propagation of electromagnetic waves, e.g. for determining the most suitable site for a radio transmitter. Instead of setting up a fixed receiver at a point  $B$  and varying the position  $A$  of the test transmitter, it was merely necessary to have a fixed transmitter at  $B$ , whose reception was measured at a variable point  $A$ . This was obviously a quicker and more economical procedure.

A simple optical example of reciprocity is the fact that when a luminous flux of intensity  $I$  impinges on a plane-parallel plate, the latter transmits a flux of  $aI$ , irrespective of which is the incident side of the plate.

There are, however, exceptions to reciprocity. In mechanics an example is provided by the behaviour of a gyroscope (fig. 2). Imagine the latter, viewed from  $z$ , to be rotating in the positive direction (clockwise). A couple, acting clockwise about the  $x$ -axis for a time  $t$  will cause the axis of rotation of the top to turn about the  $y$ -axis (i.e. in the  $x,z$  plane) through an angle  $\vartheta$ , in the *negative* direction, viewed from  $y$ . A positive couple of like magnitude, acting for a time  $t$  about the  $y$ -axis causes a turning of the top's axis about the  $x$ -axis (in the  $y,z$  plane) through the same angle  $\vartheta$ , but in a *positive* direction, viewed from  $x$ . The magnitude of the effect is in either case the same, but its sign is different.

<sup>1)</sup> Propounded by Helmholtz in *J. Math.* 56, 29, 1859; formulated for mechanics by E. Betti, *Nuovo Cimento* 1872. See also Rayleigh, *Proc. London Math. Soc.* 4, 357-368, 1873; *Sci. Papers I*, pp. 170-181 and *Enz. Math. Wiss.*, IV (4), p. 86, note.

An optical example, in which the principle of reciprocity does not apply, is Rayleigh's well-known experiment.<sup>2)</sup> This is shown diagrammatically in fig. 3. At *A* there is a polarizer (Nicol prism or polaroid sheet) mounted with its direction of polarization vertical; at *B* there is a similar polarizer, whose direction of polarization, viewed from *A*, is positioned 45° clockwise with respect to *A*. Between *A* and *B* there is a solenoid *C*, which sets up a magnetic field. In the field is a transparent rod (or column of liquid) of a substance which shows the phenomenon of magnetic rotation of the plane of polarization (Faraday effect). For diamagnetic substances this rotation is "positive", i.e. regardless of the direction of propagation, it is in the same sense as the current producing the magnetic field. (With paramagnetic substances, there is both a para-

magnetic and a diamagnetic contribution, which are usually opposite in sign). The nature of the substance, the direction of the current, the length of the coil and the strength of the field are so chosen in the experiment, that for light propagated from *A* to *B*, the rotation is clockwise and amounts to 45°. An incident beam of light from the left is therefore polarized at *A* and, on arriving at *B*, passes through completely. On the other hand an incident beam coming from right will be polarized by *B* and its direction of polarization will then be rotated anti-clockwise through 45° and will thus be perpendicular to the polarizing direction of *A*. Such a beam will be completely blocked by *A*. The system will transmit light in the *AB* direction, but not in the *BA* direction. On the other hand this non-reciprocity would not be observed if there were a naturally rotating substance (quartz, sugar solution) between the two polarizers instead of a magnetically rotating substance. In this case, if the rotation from *A* to *B* were clockwise, that from *B* to *A* would also be clockwise and the light would be transmitted in both directions.

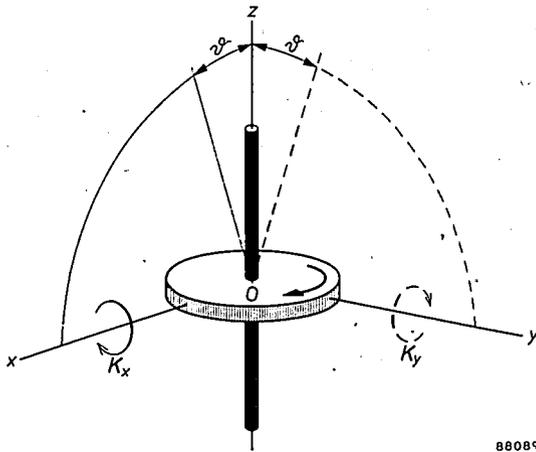


Fig. 2. A gyroscope as an example from mechanics in which the principle of reciprocity is not complied with. A couple  $K_x$  acting about the *x*-axis, causes the axis of the gyroscope to incline in the *x-z* plane. The angle  $\theta$  is proportional to the product of  $K_x$  and the time *t* for which the couple acts. Viewed from *y*, the axis of the top inclines to the left. A couple of the same magnitude and the same sense acting about the *y*-axis causes the axis of the gyroscope to incline in the *y-z* plane through an angle  $\theta$ : viewed from *x* the inclination is to the right.

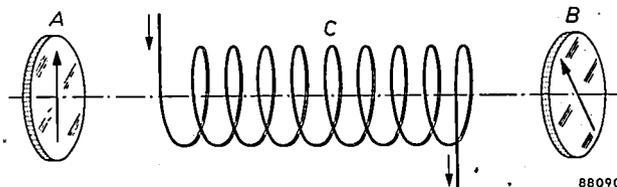


Fig. 3. Rayleigh's experiment, an optical example in which the principle of reciprocity is not obeyed. *A* and *B* are polarizers (e.g. Nicol prisms). Viewed from *A*, *B* is oriented 45° clockwise with respect to *A*. *C* is a coil containing a diamagnetic material (with a positive Faraday effect); the plane of polarization of a beam of light passing through this material is always rotated in the direction of the current producing the magnetic field. The system allows a beam of light to pass from left to right but not in the reverse direction.

Likewise in the propagation of radio waves the principle of reciprocity may be found not to apply, again in those cases in which there is magnetic rotation of the plane of polarization. The above-mentioned reciprocity of radio transmitter and receiver, for example, no longer obtains if reception is principally afforded by reflections at the Heaviside layer, and there is a noticeable rotation of the plane of polarization, brought about by the earth's magnetic field.

If it is required to construct a linear electric network that does not comply with the principle of reciprocity, it is necessary to incorporate in it the non-reciprocal network element, which Tellegen<sup>3)</sup> has called a *gyrator*, or a component whose equivalent circuit contains a *gyrator*, in addition to reciprocal elements, such as resistances, inductances and capacitances.

Such non-reciprocal components can actually be made for use with microwaves, one of the effects employed in their construction being magnetic rotation of the plane of polarization. Non-reciprocal components of this kind are made possible by the existence of ceramic ferromagnetic materials, which owing to their low conductivity are transparent to electromagnetic waves in the microwave region. Ferroxcube is such a material.

At first, ferroxcube was mainly employed for coil cores for use in the region below 10–100

<sup>2)</sup> Rayleigh, Phil. Trans. 176, 343-366, 1885; Sci. Papers II, 360-383.

<sup>3)</sup> B. D. H. Tellegen, The gyrator, an electric network element, Philips tech. Rev. 18, 120-124, 1956/57 (No. 4/5).

Mc/s, as the magnetic losses are low in this range of frequencies. In the microwave range (1000—30 000 Mc/s); in which unpolarized ferroxcube gives appreciable losses, waveguides can be constructed using magnetically polarized ferroxcube which function as non-reciprocal elements. The lower limit of frequency is then determined by the minimum magnetic field required for magnetic saturation, the upper limit by the maximum magnetic field that can in practice be produced in a waveguide.

The non-reciprocal behaviour of such elements manifests itself in the fact that the attenuation is different in opposite directions. This is of direct practical importance in microwave technique. Such elements (directional isolators, "unilines") are already being applied on a considerable scale and frequent articles on them appear in the literature<sup>4)</sup>.

In the following, a few examples of unidirectional waveguides of this type will be described, but there will first be a brief exposition of the phenomenon of gyromagnetic resonance, the working principle of such waveguides. For those interested, the theory is treated somewhat more fully in an appendix at the end of this article.

**Gyromagnetic resonance**

The mechanical example of non-reciprocity given above was that of a gyroscope. In actual fact, the gyroscopic effect is involved in the other examples, too, for electrons, owing to their orbital and their spin, show a similar behaviour. More especially, the applications of ferroxcube mentioned above are based upon the phenomenon of gyromagnetic resonance which was predicted in 1935 by Landau and Lifshitz and has been discussed in this Review<sup>5)</sup>. Landau and Lifshitz demonstrated that a spinning electron, oriented by a magnetic field *H* executes a free precession about the direction of this field, with angular velocity

$$\omega_p = \gamma H, \dots \dots \dots (1)$$

where<sup>6)</sup>

$$\gamma = 2.21 \times 10^5 \text{ m/A.sec.} \dots \dots \dots (2)$$

The direction of rotation of the precession is the same as that of a current which would produce the

field *H*. An alternating magnetic field, perpendicular to *H*, gives rise to a forced precession with the same direction of rotation as the free precession and with the angular frequency  $\omega$  of the alternating field. If  $\omega$  differs little from  $\omega_p$ , resonance occurs and the precession becomes very large.

What has been said above holds primarily for a free electron, but *mutatis mutandis* it also applies to an electron in a ferromagnetic crystal. In a crystal the free precession is damped and the forced precession is attended with losses, which become particularly great when  $\omega \approx \omega_p$ . Further, the value of  $\omega_p$  is in general no longer determined by the simple formula (1): it depends in a more complex manner upon the magnetic state of the material. For an ellipsoid of ferroxcube with one of its axes parallel to the external field *H<sub>e</sub>* (*z* direction),  $\omega_p$  follows from Kittel's formula (see articles I and II, footnote 5)):

$$\omega_p = \gamma \sqrt{[H_e + (N_x - N_z)M_s] [H_e + (N_y - N_z)M_s]}, \dots \dots (3)$$

where *M<sub>s</sub>* is the saturation magnetization and *N<sub>x</sub>*, *N<sub>y</sub>* and *N<sub>z</sub>* the demagnetization factors (*N<sub>x</sub>* + *N<sub>y</sub>* + *N<sub>z</sub>* = 1).

In view of what is to follow, it is useful to regard the alternating magnetic field, which is applied perpendicularly to *H<sub>e</sub>* and causes the forced precession, as a superposition of two opposing "rotating fields" (each of which may be represented as a vector that rotates in a plane perpendicular to *H<sub>e</sub>*). Only the field that rotates in the same direction as the free precession, leads to resonance and its associated losses. The opposing rotating field has scarcely any effect on the spins.

A direct consequence of this is that the Faraday effect can be observed with plane polarized magnetic waves propagated in the ferroxcube in the direction of the magnetizing field *H<sub>e</sub>*. The resolution of the alternating magnetic field of the wave into two opposing rotating fields gives rise to two circularly opposed polarized waves. On account of the difference mentioned above in their interaction with the spins, these waves have different rates of propagation.

<sup>6)</sup> For  $\gamma$  we have the relationship

$$\gamma = g \frac{|e|}{2m_0} \mu_0$$

Here,  $|e|$  is the absolute value of the charge on the electron ( $1.6 \times 10^{-19}$  coulomb),  $m_0$ , its rest mass ( $0.9 \times 10^{-30}$  kg),  $\mu_0 = 4\pi \times 10^{-7}$  Vs/Am and *g* the Landé factor. For a free electron the value of the latter is approximately 2, for an electron bound in a ferromagnetic crystal lattice, its value may vary somewhat, viz. from 2.00 to 2.25, owing to a correction for the orbital field.

In the literature the symbol is sometimes used for the negative quantity  $g(e/2m_0)\mu_0 = -g(|e|/2m_0)\mu_0$ .

<sup>4)</sup> C. L. Hogan, Ferromagnetic Faraday effect, Bell Syst. tech. J. 31, 1-31, 1952. A. G. Fox, S. E. Miller and M. T. Weiss, Behaviour and applications of ferrites in the microwave region, Bell Syst. tech. J. 34, 5-103, 1955.  
<sup>5)</sup> H. G. Beljers and J. L. Snoek, Gyromagnetic phenomena in ferrites, Philips tech. Rev. 11, 313-322, 1949/50 hereafter to be referred to as I; H. G. Beljers, Amplitude modulation of centimetre waves by means of ferroxcube, Philips tech. Rev. 18, 82-86, 1956/57, hereafter to be referred to as II.

Owing to the phase difference between the two circular components, which this brings about, the plane polarized wave displays a rotation of its plane of polarization which is proportional to the distance covered in the ferroxcube.

**Non-reciprocal elements employing Faraday rotation**

For the transmission of microwaves rectangular waveguides are normally used, in which the long side  $a$  of their cross-section satisfies:  $a < \lambda < 2a$ . The propagation in the waveguide is then brought about by a wave whose mode of vibration is designated  $TE_{10}$ , the electric vector being parallel to the short side of the rectangle (see fig. 4a).

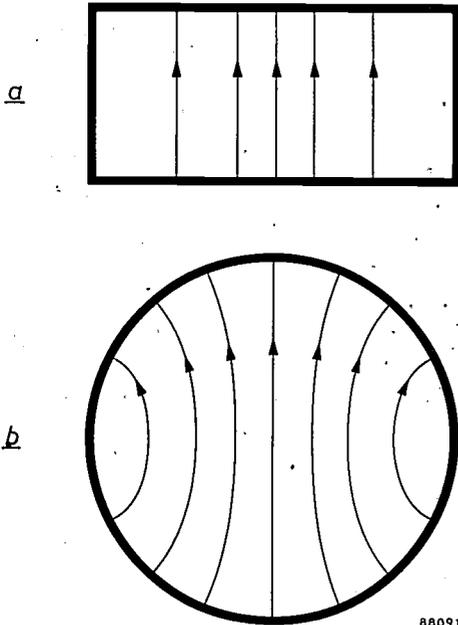


Fig. 4. a) Rectangular waveguide with mode of vibration  $TE_{10}$ ; b) Circular waveguide with mode of vibration  $TE_{11}$ . The electric lines of force are indicated in both cases.

For an element with rotation of the plane of polarization it is customary to employ a circular waveguide with propagation in the  $TE_{11}$  mode (fig. 4b). The ferroxcube usually takes the form of a thin cylindrical rod mounted coaxially in the waveguide and magnetized along its axis (if the waveguide were to be completely filled with ferroxcube, a large number of modes of vibration could occur simultaneously, and this would lead to undesirable effects, especially with long rods). An example of the rotation obtained with the above-described arrangements at a frequency of 24 000 Mc/s, as a function of the axial magnetic field may be seen in fig. 5. The angle of rotation increases with the increasing strength of the magnetic field, until saturation is reached.

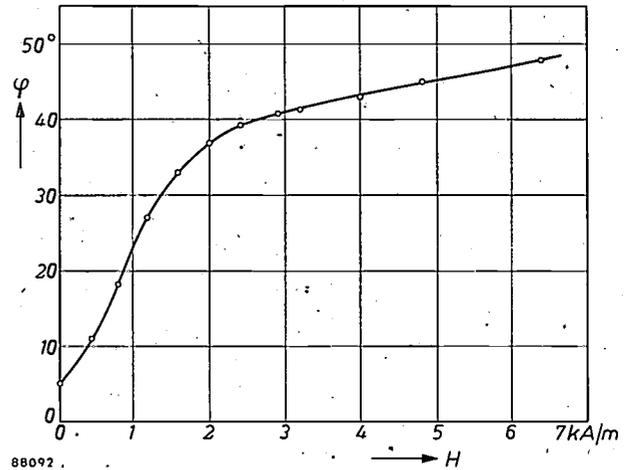


Fig. 5. The rotation  $\phi$  of the plane of polarization of an electromagnetic wave at the end of a circular waveguide fitted with an axial rod of magnetized ferroxcube (length 4.4 cm), as a function of the strength  $H$  of the magnetizing field.

The essential condition now is that the sign of the rotation should depend upon the direction of the magnetic field with respect to the direction of propagation. The circular waveguide, provided with a ferroxcube rod, can be incorporated in a system of rectangular waveguides (see fig. 6) with the aid of adapters (the cross-sections of which change gradually from rectangular to round to avoid reflection). If the adapters at either end are at  $45^\circ$  with respect to each other, and if the Faraday rotation is also set at  $45^\circ$ , the wave will be transmitted almost completely in one direction. In the reverse direction the wave cannot travel further after rotation, since the electric vector is now parallel to the long side of the rectangular section, which does not represent a possible mode of vibration. It can be arranged that the energy is not reflected, but absorbed on the spot, by mounting a thin plate of a semi-conductor e.g. graphite-containing "Philit", parallel to the electrical lines of force.

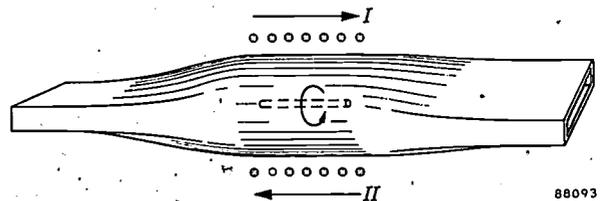
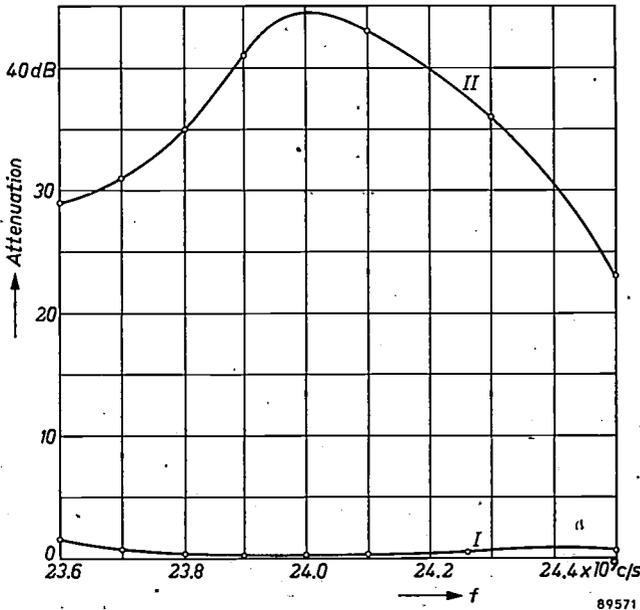


Fig. 6. Construction of a directional isolator, based on the Faraday effect. If the magnetic field be applied in the direction  $I$ , a wave can then be propagated through the waveguide only in that direction.

Such a system is a directional isolator, because if it be interposed between two equipments  $A$  and  $B$ ,  $A \rightarrow B$  being the desired direction of transmission, it will prevent any waves that might be reflected at  $B$  from returning to  $A$ ;  $A$  is thus "isolated" from

these undesirable waves from *B*. Since the Faraday rotation depends upon the frequency, the element can be employed only in a limited range of frequencies (for a given magnetic field). *Fig. 7* gives the measured values of the attenuation in both directions in the frequency range 23 600—24 500 Mc/s for such a directional isolator constructed in this laboratory.

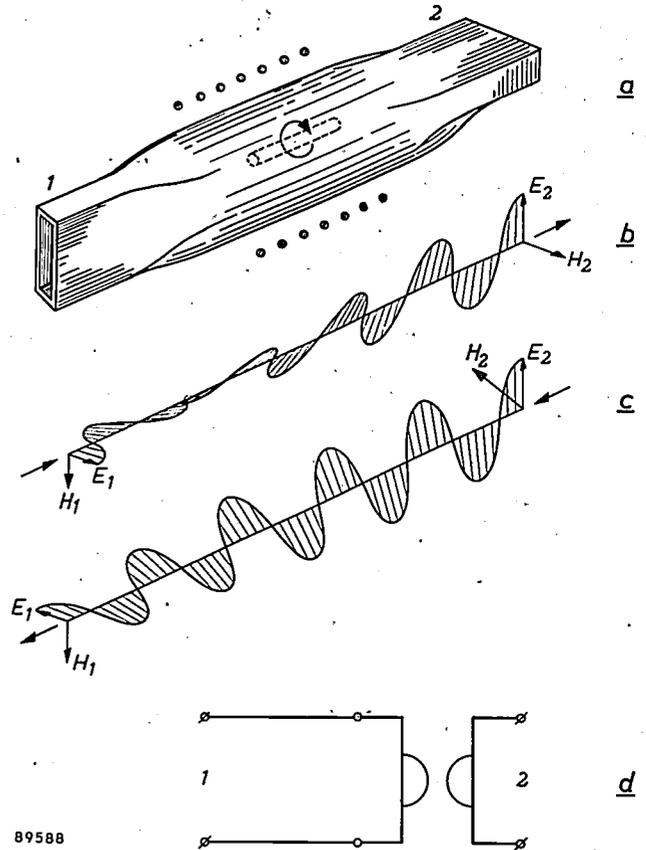
The polarizing magnetic field at the axis of this isolator is about 5000 A/m (60 Oe). It is produced by a few rings of axially magnetized ferroxdure<sup>7)</sup>, which can be slipped round the waveguide. Some measure of control can be exercised over the strength of the field by moving the rings along the waveguide and the angle of rotation can thus be adjusted to the correct value. Instead of a ferroxcube rod, a rod of ferroxdure, magnetized along its longitudinal axis can be mounted coaxially in the waveguide. An external magnetic field is then unnecessary. Care must be taken when using this directional isolator to avoid working under such conditions that the circular component, which has the same direction of rotation as the spin precession,



*Fig. 7.* Attenuation brought about by a directional isolator of the type shown in *fig. 6*, as a function of the frequency *f*, in the frequency range 23 600—24 500 Mc/s. Curve I is for the forward direction in which the wave is transmitted, curve II for the reverse direction in which the wave is suppressed.

is too strongly damped. This would give rise to inequality of the two circular components, and hence would diminish the attenuation action. This is equivalent to saying that the wave frequency should

not approximate too closely to the resonance frequency  $\omega_p$ . The axial magnetic field must not, therefore, be given too high a value. A strong field is in fact not required to magnetize a thin rod.



*Fig. 8. a)* Waveguide showing a difference in transmission length of  $\lambda/2$  in the two directions (after Hogan, loc. cit.<sup>4)</sup>). *b)* Instantaneous view of wave propagated in one direction. *c)* Instantaneous view of wave propagated in the opposite direction to *b)*. *d)* Equivalent circuit of the waveguide, comprising a Lecher line and an ideal gyrator<sup>8)</sup>.

The waveguide shown can be seen to possess the properties of a gyrator from the relationship between the electric vector *E* of the incident wave and the magnetic vector *H* of the emergent wave. In the one case *b)* we have  $H_2 = E_1/Z$ , whilst in the other case *c)*  $H_1 = -E_2/Z$ , *Z* representing the characteristic impedance. These relations (or their transmission line equivalents,  $i_2 = v_1/Z$  and  $i_1 = -v_2/Z$ ) may be compared with the formulae for the gyrator (see eq. (9) in the article cited in note<sup>3)</sup>).

Such an element can also be arranged to give a rotation of 90° instead of 45°. This will result in equal transmission in both directions: the non-reciprocal behaviour then manifests itself in the fact that the phase differs by 180° for the two directions. Such an element is an “ideal gyrator” as postulated by Tellegen. This is illustrated in *fig. 8*.

<sup>8)</sup> For equivalent circuits of other unidirectional waveguides, see H. J. Carlin, Principles of gyrator networks, Proc. Symp. Mod. Adv. Microwave Techn. New York 1954, p. 175.

<sup>7)</sup> See Philips tech. Rev. 14, 194-208, 1952/53.

**Non-reciprocal elements with transverse magnetic fields**

Propagation in a transverse magnetic field can also be made use of in the construction of non-reciprocal elements. Here a normal rectangular waveguide (see fig. 4a) is used, in which the propagation is in the TE<sub>10</sub> mode. Consider a system of coordinates with the z-axis parallel to the short side of the rectangle, and the x-axis parallel to the long side of the rectangle; the y-axis then coincides with the longitudinal axis of the waveguide, i.e. the direction of propagation. The electrical lines of force are parallel to the z axis; the magnetic lines of force are closed loops in the x,y plane as shown in fig. 9a<sup>9)</sup>.

The magnetic field has components:

$$\left. \begin{aligned} h_x &= j\beta A \cos(\pi x/a) \exp[j(\omega t - \beta y)] \\ h_y &= (\pi/a)A \sin(\pi x/a) \exp[j(\omega t - \beta y)] \\ h_z &= 0. \end{aligned} \right\} \dots (4)$$

where *A* is an arbitrary constant, *a* the length of the longer side of the waveguide and  $\beta$  the propagation constant in the waveguide ( $\beta = 2\pi/\lambda_g$ ). Considering a point *x, y* we see that the magnitude and direction of the magnetic vector at that point vary periodically. For certain values of *x*, however, the nature of this variation is such, that only the direction of the magnetic field changes and not its strength, so that the magnetic vector describes a circle and a local rotating field is set up. This is the case for those values of *x*, for which  $h_x = \pm jh_y$  (equal amplitude, phase difference 90°) or:

$$\tan \frac{\pi x}{a} = \pm \frac{\beta a}{\pi} = \pm \frac{2a}{\lambda_g} \dots (5)$$

Two values are found,  $\pm x_1$ , on either side of the centre. For one of these values, the rotation of the magnetic vector, viewed along *z*, is positive, for the other negative. Both rotations change sign as the direction of propagation changes sign.

If, now, a thin plate of ferroxcube be mounted in the waveguide at *x*<sub>1</sub>, and a constant magnetic field *H* be applied in the *z*-direction, i.e. across the direction of propagation of the wave, the situation is the same as that described earlier (page 160): a rotating magnetic field *h* perpendicular to the constant field *H*. For that direction of propagation in the waveguide for which the direction of rotation of *h* in the ferroxcube plate is the same as that of the spin precession around *H*, gyromagnetic resonance can occur, with corresponding damping of the wave. A wave with the reverse direction of propagation passes through almost unimpeded.

Unlike the previous case (Faraday effect), in which the damping must remain small, it is precisely on the damping in the material that the working principle is based, and the damping must therefore

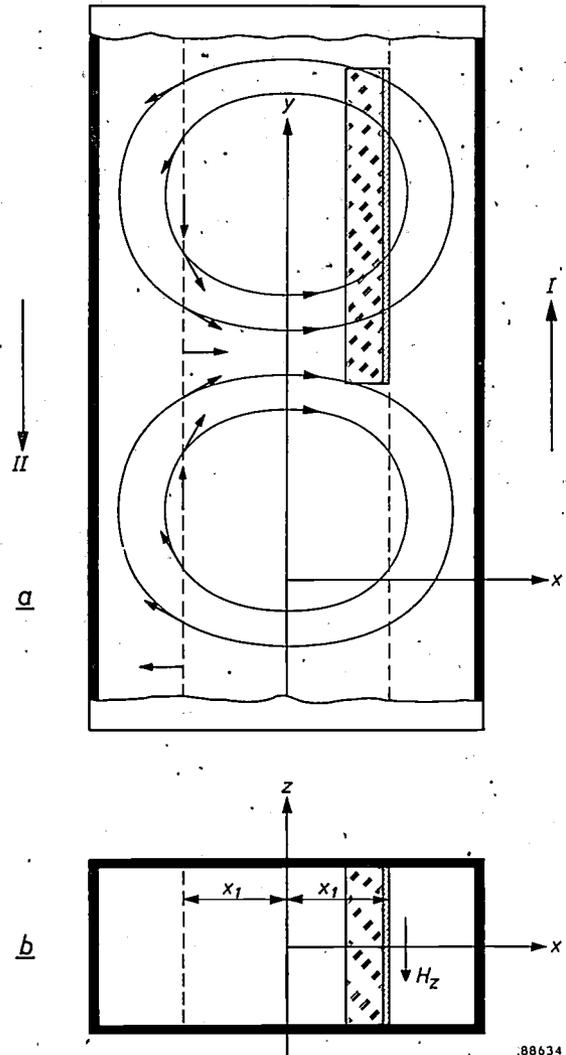


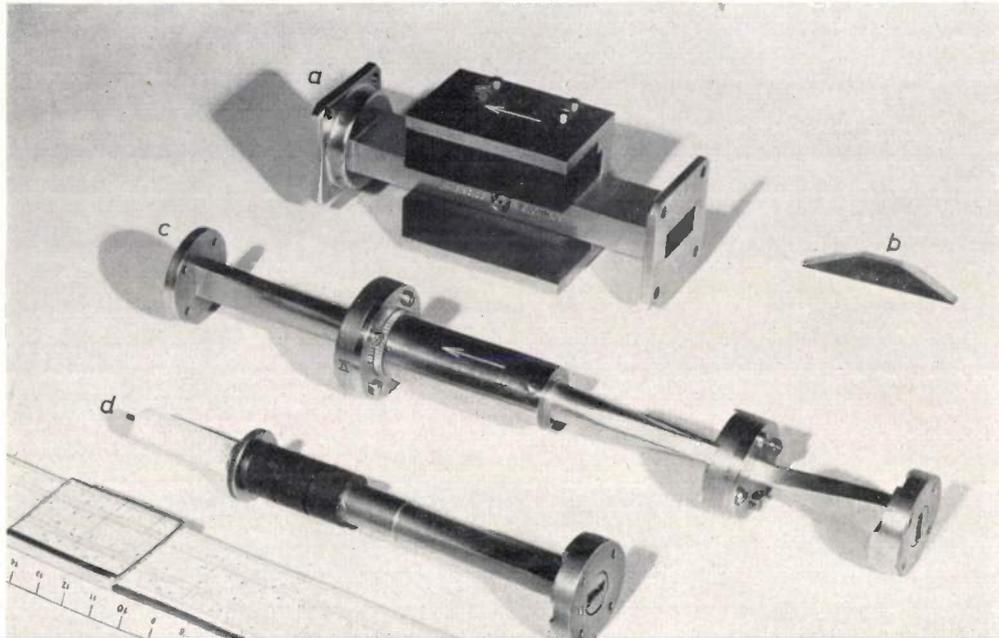
Fig. 9. The magnetic lines of force in a rectangular waveguide with mode of vibration TE<sub>10</sub>. At points at a certain distance *x*<sub>1</sub>, on either side of the axis, the magnetic vectors are equal and constant in magnitude and rotate uniformly. b) Section of the waveguide. By mounting a ferroxcube plate magnetized in the *z*-direction (across the direction of propagation) in the *yz* plane and distant *x*<sub>1</sub> from the axis, gyromagnetic resonance is set up in the plate and the wave is therefore attenuated in direction *II*.

be made as great as possible. This means that  $\omega$  must lie close to  $\omega_p$ . Although the presence of the plate, which has a fairly high dielectric constant ( $\epsilon_r = 10$ ) and sometimes also a fairly high permeability, strongly distorts the original field, it is nevertheless possible to mount the plate in such a way that the wave whose magnetic vector in the plate rotates in the opposite sense to the spin precession passes through almost unattenuated.

<sup>9)</sup> M. L. Kales, H. N. Chait and N. G. Sakiotes, J. appl. Phys. 24, 816-817, 1953.

On closer examination it is found that the correct position for the plate is not exactly at  $x_1$ , since then the purely circular polarization of the magnetic vector occurs in the space outside the plate but not in the ferrite material itself, owing to the demagnetization in the plate. The plate must therefore be shifted somewhat nearer to the centre of the waveguide<sup>10)</sup>. The resonance frequency can be most simply derived from Kittel's formula (3), by

or direction in all parts of the plate. For this and a number of other reasons, such as magnetic anisotropy in the crystals, porosity and internal stresses, the value of  $\omega_p$  in the material is not accurately determined, and there is no sharp resonance. This is to be regarded as an advantage, as it leaves a certain freedom in the frequency, and one-way transmission systems so constructed can be used for a fairly wide range of frequencies.









in mercury vapour under low pressure with additions of noble gases had been known since about 1910 (Claude and other). These two innovations made it possible to ignite the high-pressure mercury vapour lamp without the tilting operation and, with suitable geometry (combined where necessary with auxiliary electrodes or other artifice), the lamp could be made to ignite from a mains voltage of 220 V. In this way it became practicable to use the lamp for lighting purposes, especially the lighting of streets, open spaces, factory sheds and the like, where there was a need for a light source with very high efficiency (incandescent lamps can reach only about 15 lm/W) and where, at that time, the poor colour rendering of its light was not regarded as a serious objection.

An important feature of the new lamps was that the amount of mercury introduced into the envelope was limited to that which would just be completely vaporized once the desired pressure had been reached, the vaporization taking place as the lamp warmed up. This prevented the vapour pressure rising to an excessive value should the mains voltage or ambient temperature become too high.

The first high-pressure mercury vapour lamps for lighting purposes were made of hard glass with a high softening point (of the order of 700 °C) and not of silica as used in the lamp developed by Küch and Retschinsky. The reason was not merely the high price of silica, but also the problem of the connections through the wall of the tube, a basic and recurring problem in the history of this type of lamp. The method used by Küch and Retschinsky, viz. ground conical holes in which the terminals made a tight fit, was not suitable for a commercial product and the fusing of wires into silica was still a matter of difficulty in the 1930's.

The power of one type of these high pressure mercury lamps (type HO) [MA/V]<sup>3)</sup>, which were supplied from the mains via a choke coil, amounted to about 400 W and the pressure inside them to about 1 atm; the actual discharge tube was enclosed in an outer envelope. In this way the lamp was made easier to handle and, when used in the open air, the discharge tube was protected against wind, so that the warming-up process and the building-up of the mercury vapour pressure were not interfered with.

In the further development of the high-pressure mercury discharge the laboratories of the various lamp factories took diverging paths. One point

was of great importance for the line of development followed by Philips: the appreciation of the fact<sup>4)</sup> that the luminous efficiency  $\eta$  of a high-pressure mercury discharge, is dependent in the first instance on  $P$ , the power loading per cm of arc length, and only slightly dependent on the tube diameter and the vapour pressure once the latter has exceeded a certain value.  $\eta$  increases with  $P$  in the manner shown in fig. 3.

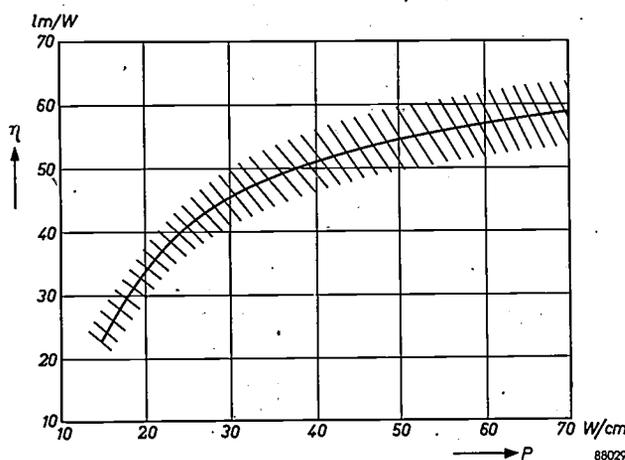


Fig. 3. Luminous efficiency  $\eta$  of a mercury vapour discharge, as function of the power load  $P$  per cm (schematic). With sufficiently high vapour pressures — higher as the tube diameter is made smaller — the observed values for different tube diameters and pressures lie within the hatched area.

This behaviour is a consequence of the contraction of the discharge. Apart from the pressure, which remains constant throughout the tube, it is the temperature that determines excitation and ionization in the high-pressure discharge, in accordance with the equations of Boltzmann and Saha (see<sup>4)</sup> or<sup>2)</sup>). The temperature decreases from the axis of the tube towards the wall, owing to dissipation of heat to the surroundings. For these reasons the discharge mainly takes place in the central portion of the tube, where the temperature is of the order of 6000 °K. As a result of this contraction of the discharge, the losses are largely restricted to heat conduction losses, these being determined by the difference of temperature between the centre and the wall. If we call the heat conduction losses per cm length  $A$ , then a power  $P - A$  per cm of tube length leaves the discharge in the form of radiation. Of this quantity a certain fraction  $g$  is absorbed in the mercury vapour between the contracted discharge and the wall, in the tube wall itself and in the outside air (which absorbs, for example, the ozone-forming radiation

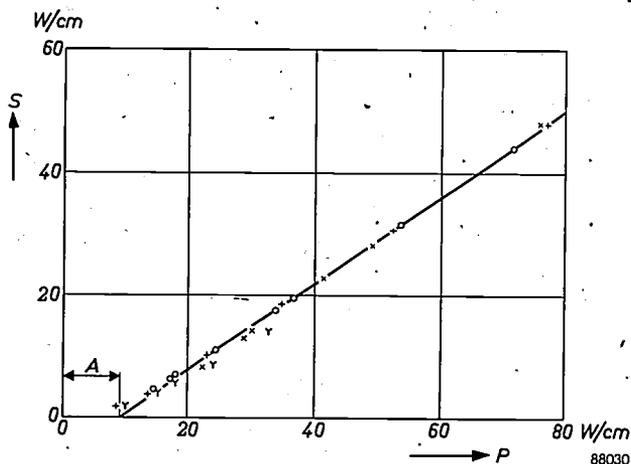
<sup>3)</sup> In this article, lamp types are designated first by the Philips code letters and then, in square brackets, by the code letters used in British territories and certain other countries.

<sup>4)</sup> W. Elenbaas, Die Quecksilber-Hochdruckentladung, Physica 1, 673-688, 1934.

of wavelength of 1850 Å). Hence  $S$ , the measured radiation per cm of tube length, will be

$$S = (1 - g) (P - A) \dots \dots \dots (1)$$

In *fig. 4*,  $S$  is plotted against  $P$  for a series of high-pressure mercury discharges in silica tubes. It will be seen that the points fit equation (1) very closely,



*Fig. 4.* Total radiation  $S$  per cm of arc length from the high-pressure mercury vapour discharge, as function of  $P$ , the power load per cm, measured on a series of silica discharge lamps. The tube diameter was 2.7 cm in all cases. The different points (+, 0, etc) in the graph indicate a sub-series of measurements on tubes differing in the quantity of mercury they contained per cm of their length. Similar series of measurements were performed on tubes with diameters of 0.92 cm and 0.33 cm, these producing roughly the same curve.

the heat conduction loss  $A = 10$  W/cm being fairly constant (i.e. independent of the power  $P$ , the tube diameter and the pressure) and the factor  $1-g$  being 0.72. Consequently  $\eta$  the luminous efficiency, which is roughly speaking proportional to  $S/P$ , is also to a first approximation proportional to  $(1-g) (1-A/P)$ ; hence, as  $P$  increases,  $\eta$  rises first rapidly and then less steeply. This agrees well with the curve obtained by measurement in *fig. 3*.

The following may serve to show that  $A$ , the heat conduction loss from the contracted discharge per cm of tube length is largely independent of the power  $P$ , the tube diameter and the pressure. If  $T$  is the temperature and  $\lambda$  the thermal conductivity, then the heat conducted away per second through an imaginary cylinder at a distance  $r$  from the axis will be:

$$A = -2\pi r \lambda \frac{dT}{dr} \dots \dots \dots (2)$$

The temperature of the discharge itself varies very little from one discharge to another, since  $T$  occurs only as an exponent of  $e$  in the Boltzman equation: hence a small variation in  $T$  corresponds to a large variation of the radiation. As a consequence, the quantity  $r dT/dr$  is practically the same at corresponding points in different discharges (in tubes of diameters  $a_1, a_2$ , for example, points  $r_1$  and  $r_2$  are corresponding if  $r_1/a_1 = r_2/a_2$ ), since the wall temperature may be taken to be approximately the same in all cases. Further,  $\lambda$  is independent of the pressure and also independent of  $P$ , since the latter can only exercise an influence on  $\lambda$  through  $T$  which, as we have seen, varies only slightly from one discharge to another. Hence  $A$  is substantially independent of the power  $P$ , the tube diameter and the pressure.

From this relationship between  $\eta$  and  $P$  it followed that, for high luminous efficiency, it was advantageous to have a short arc for a given amount of power. It was also clear that very high brightnesses might be attained by increasing the power  $P$  per cm of arc length, for — assuming that the diameter of the discharge remains the same — the power radiated per unit of surface increases more than proportionally to  $P$ , owing to the increase in luminous efficiency. This opened up a prospect of new applications inaccessible to the incandescent lamp whose brightness is limited to about 2000 cd/cm<sup>2</sup> by the melting point of tungsten and whose lifetime is then only some hours or tens of hours.

At first it was attempted to obtain a high value of  $P$  by arranging for the discharge to take place in the centre of a spherical hard glass vessel. However, this turned out to be impracticable in that, with the high power involved, the bulb had to be large in diameter in order to keep the wall temperature within bounds; but then the vapour pressure had to be lower, on account of the danger that the relatively large bulb might explode. This meant a decrease in the contraction of the discharge, and the desired gain in brightness was only partially realised. Moreover, at lower vapour pressures the gradient (voltage drop per cm of discharge length) is smaller, so that in order to develop a large power per cm the current has to be made much heavier; this in its turn created difficulties with regard to the electrodes and seals.

The history of the lamp then took an unexpected turn. In the Philips Laboratory at Eindhoven tungsten wires were successfully sealed in silica<sup>5)</sup> with the aid of an intermediate glass, thus giving seals suitable for moderate currents. A small high-pressure mercury vapour lamp was designed on this basis, the discharge taking place in a silica tube of capillary form<sup>6)</sup> (see *fig. 5*). The combined problem of wall temperature, tube diameter, pressure and danger of explosion was avoided with this design, inasmuch as the silica tube could be water-cooled. The resulting design was ideal for attaining high degrees of brightness. The pressure of mercury vapour in the capillary can be allowed to become very high indeed (of the order of 100 atm.) and this makes the discharge path, already narrow, even more constricted. The voltage gradient thus becomes very large. In consequence, the load per cm

<sup>5)</sup> B. Jonas, The sealing of metal leads through hard glass and silica, Philips tech. Rev. 3, 119-124, 1938.

<sup>6)</sup> C. Bol, W. Elenbaas, W. de Groot, De Ingenieur 50, E91, E83 and E92, 1935.

can be made very high for a reasonable current value (about 1 A) which will not cause too much trouble with the seals and electrodes. Lamps of this type (SP) [MD] have been developed for cinema projectors amongst other purposes; brightnesses of about 50 000 cd/cm<sup>2</sup> are attained in the axis of

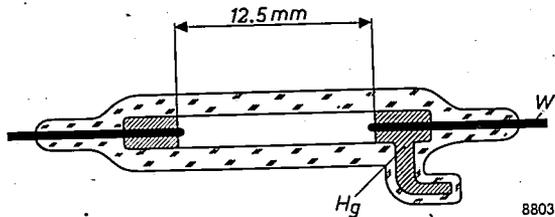


Fig. 5. Water-cooled super-pressure mercury vapour lamp of capillary form (SP type) [MD]; it is rated at a power of 1000 W for an arc length of 12½ mm, the vapour pressure being of the order of 100 atm.

the lamp, the diameter of the capillary being 2 mm. With experimental forms brightnesses of 180000 cd/cm<sup>2</sup> — greater than that of the sun<sup>7)</sup> — have been attained. A further advantage of the very high vapour pressures and current densities in super-pressure lamps is the strong continuum in the spectrum of their radiation, this resulting in light of a colour rendering that is a marked improvement on that given by high-pressure mercury vapour lamps with pressures of the order of 1 atm.

It may be mentioned that the capillary form of the original SP lamp was also of importance in regard to the design of the seals. The lamp was filled with an excess of mercury which, on account of capillarity, remained lodged as a drop at each end while the lamp was burning and which ensured a uniform and relatively low temperature at the critical regions adjoining the seals. The tungsten electrodes just protude from the drops of mercury (fig. 5). Owing to the very high pressure inside the tube, the electrodes evaporated only slowly in spite of their high temperature.

Later there was a change to another, simpler, form of lead-in through silica, invented by Gabor<sup>8)</sup>. A molybdenum ribbon, less than 20 μ thick and a few mm wide, is fused directly into the silica. The ribbon is sufficiently ductile to accommodate the difference in the expansions of molybdenum and quartz glass. This method later allowed the making of water-cooled capillary lamps containing a precise quantity of mercury which was vaporized

almost completely whilst the lamp was burning<sup>9)</sup>.

Fairly heavy currents can be passed through the silica wall by fusing in a number of molybdenum ribbons side by side. On this basis Rompe and Thouret<sup>10)</sup> constructed spherical silica lamps of very high brightness; this, then, is the idea mentioned on the previous page, which was difficult to realize in hard glass. In the war years these lamps, known as "compact source" lamps (CS lamps) [ME], were further developed in Germany and England for searchlights. They are still being made today in various power ratings and arc lengths for a number of special purposes.

Let us now return to the development of high-pressure mercury vapour lamps for general lighting purposes. The hard glass 400 W lamps, with vapour pressures of about 1 atm., as described earlier, had to some extent been adopted for public lighting installations. There was a demand for smaller units: 250 W and even 150 W lamps appeared, but for these wattages the advantage over the incandescent lamp in the matter of efficiency was less marked. The reason for this is that if, in view of the lower power, the arc-length is shortened in order to step up  $P$  and hence also  $\eta$ , in accordance with fig. 3, the total arc voltage becomes rather low in comparison with the mains voltage and there are excessive losses in the ballast and at the electrodes. One cannot avoid this by raising the voltage per cm of arc-length, since the reduction in diameter and/or increase of pressure that this would necessitate would mean high wall temperatures and the danger of explosion. One must therefore be content with a relatively long arc, a lower  $P$  and a lower  $\eta$ .

The problem was then attacked from a different point of view, basing design considerations on the super high pressure lamp. The use of a fused silica envelope for the latter made water-cooling possible by virtue of the low coefficient of expansion. This kind of cooling cannot of course be considered for general lighting purposes. Even without forced cooling, however, silica has the advantage over hard glass that its softening temperature is considerably higher. It was thus possible to take the SP lamp, omit the water-cooling, and increase the internal diameter of the silica capillary slightly (to 4 mm), thereby dissipating a  $P$  of 40 - 50 W/cm (in the super-pressure lamps 400 - 1000 W/cm were dissipated). A lamp of this type<sup>11)</sup> (HP 300) [MB/U 300]

<sup>7)</sup> W. Elenbaas, Z. techn. Physik 17, 61, 1936; see also Philips tech. Rev. 1, 62, 1936.  
<sup>8)</sup> D. Gabor, D.R.P. 573 448, 1931.

<sup>9)</sup> J. Kern, Z. techn. Physik 20, 250, 1939.

<sup>10)</sup> R. Rompe en W. Thouret, Z. techn. Physik 17, 377, 1936.

<sup>11)</sup> Philips tech. Rev. 1, 129, 1936.

having an arc of length 18 mm and a mercury vapour pressure of about 20 atm. consumed 75 W and gave a luminous flux of 3000 lumen, so that its efficiency was 40 lm/W. The arc voltage was 230 V and the lamp was supplied from a transformer with an open-circuit voltage of 410 V. The tungsten lead-in wires were fused into the silica with various kinds of intermediate glasses. The silica tube was accommodated in an outer bulb having the same shape as an incandescent lamp.

In the further development of these lamps even more water was put into the wine by using the same principle for lamps to operate on an open-circuit voltage of 220 V. For this purpose the arc voltage had necessarily to fall to 120 V, and to this end the diameter was again increased somewhat and the pressure lowered. Moreover, the arc had to be made rather longer, because the blackening at the tube ends arising during the lifetime of the lamp caused too much loss of luminous flux. Nevertheless, as regards power load per cm and efficiency, these high pressure lamps were considerably superior to the older ones made of hard glass, so much superior in fact that even the larger types of the old lamp were largely supplanted by the high pressure lamps now developed for higher power ratings (125 W, 250 W, 400 W, and 1000 W).

The small dimensions of the high pressure quartz lamps did much to open up the terrain of interior lighting for the high-pressure mercury vapour lamp. Hard glass mercury vapour lamps using an incandescent filament<sup>12)</sup> as stabilising element in place of the choke had already been used in England. The high power of the mercury vapour lamp by itself produced units that were too big for practical purposes but by adding a series-connected incandescent filament to the smaller quartz lamp accommodated in the existing outer envelope, a lamp (type ML) [MBT/U] of convenient size was produced which can be connected to the mains without a ballast device<sup>13)</sup>. On account of the low efficiency of the incandescent filament, the total efficiency drops to about 20 lm/W: however the incandescent filament makes a useful contribution to the light emitted and the lamp as a whole gives quite a satisfactory colour rendering, and can therefore be used for

offices, public halls, and other utility areas requiring a high illumination level.

A further advantage offered by the quartz HP lamp was that it radiated the ultra-violet portions of the spectrum, hitherto absorbed by the walls of hard glass lamps. By giving the ML lamp just described an outer envelope that transmitted ultra-violet rays; it was possible to use it as a simple and convenient source for medical irradiation purposes (MLU lamps) [MBTR/U]. Larger units for these purposes were equipped with special lamps of the high-pressure type having no outer envelope and provided with a choke as ballast; they had power ratings of 500 and 250 W ("Biosol"). These, then, are the modern counterparts of the lamp developed by Kűch and Retschinsky fifty years ago and used for the same purpose.

Also of interest is the development whereby the ultraviolet radiation emitted by the discharge tube was exploited for the production of visible light; for this purpose, the inner walls of the outer envelope were covered with fluorescent powder (HPL lamps) [MBF]. This has improved the colour rendering to such an extent that the lamps have already largely supplanted the original HP lamps for street lighting purposes. The fluorescent substances employed in these lamps have been the object of intensive research. Recently red fluorescing arsenates and germanates have been put to use<sup>14)</sup>.

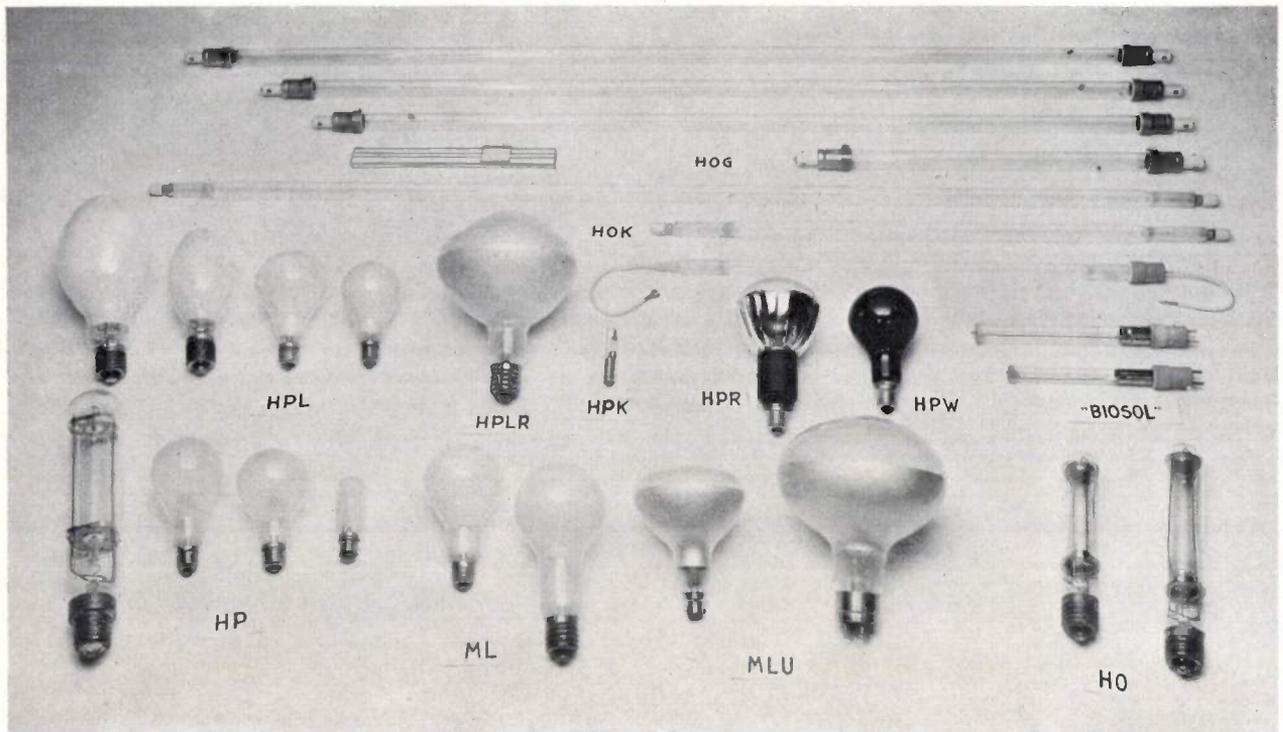
The HPW lamp [MBW/U] may be regarded as the opposite to the fluorescent lamp; in this case, the mercury discharge tube is accommodated in an outer envelope that transmits no visible light but only radiation of the near ultra-violet. These lamps are used, amongst other things, for the fluorescent detection of forgeries and for special fluorescent effects on the stage, in advertising, in shop windows, and so on.

In conclusion mention may be made of a lamp type which, lying outside the mainstream of development of the high-pressure mercury vapour lamp, exploits its elongated form and particular kind of radiation. This is the photo-printing lamp. Sensitized paper, running beneath such a lamp of length somewhat greater than the paper breadth, is exposed rapidly and evenly. If a moderate paper speed is acceptable, the load per cm need only be relatively low and hence one can revert to the use of hard glass. Versions of such lamps exist in both

<sup>12)</sup> W. H. Le Maréchal and J. N. Aldington, Brit. Pat. No. 447 428, 1936. Once the arc voltage has reached its full value, part of the incandescent filament is short-circuited by a bi-metallic relay.

<sup>13)</sup> K. Larché and E. Summerer, Licht 10, 172, 1940. Also J. Funke and P. J. Oranje, The development of blended-light lamps, Philips tech. Rev. 7, 34-40, 1942.

<sup>14)</sup> J. L. Ouweltjes, W. Elenbaas and K. R. Labberté, A new high-pressure mercury vapour lamp with fluorescent bulb, Philips tech. Rev. 13, 109-118, 1951-52.



88135

Fig. 6. A varied collection of lamps of the high-pressure mercury vapour family. In addition to the types mentioned in the text, three other lamps are shown: the HPK lamp, i.e. a high pressure discharge tube without an outer envelope, used as a source of ultra-violet radiation of very short wavelength (2537 and 1850 Å), and the HPR [MBR/U] and HPLR [MBFR] lamps, which are provided with an internal reflector.

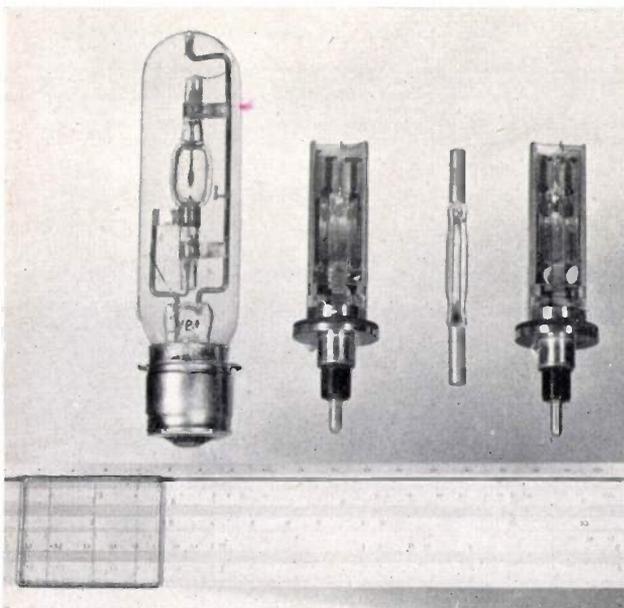


Fig. 7. Mercury vapour lamps which work under very high pressures. From left to right: a compact source lamp rated at 150 W; a 500 W super-pressure lamp (in reflector) for A.C. supply and water-cooling; a 900 W super-pressure lamp for forced air-cooling; a 1000 W super-pressure lamp for DC supply and water-cooling. The last three types are used for laboratory purposes, for cloud-height meters on airfields and for cinema projectors respectively.

hard glass (HOG) [MA/U] and quartz (HOK) [MB/U], and of various dimensions and power ratings.

It will thus be seen that the lamp invented by Küch and Retschinsky has given birth to a whole family of high-pressure mercury vapour lamps suitable for all kinds of different applications. Photographs of some the lamps are reproduced in *figs. 6 and 7*. Development continues in the direction of higher power ratings and higher brightness values and of special forms such as lamps having built-in reflectors, and so on.

**Summary.** Review of the development of different types of high-pressure mercury vapour lamps from the lamp designed by Küch and Retschinsky in 1906 and used as a "sunray lamp". The relationship between luminous efficiency (and brightness) and the power load per cm of arc-length is discussed, this being fundamental in the design of super pressure lamps with their extremely high brightness values: up to 180000 cd/cm<sup>2</sup>. There follows a discussion of the considerations that led to high pressure lamps and their derivatives, the ML [MBT/U], HPL [MBF] and HPW [MBW/U] lamps. Mention is also made of the "Biosol", the HOK [MB/U] and HOG [MA/U] photo-printing lamps and the spherical "compact source" lamps.

# A MAGNETODYNAMIC GRAMOPHONE PICK-UP

## II. FREQUENCY CHARACTERISTICS

by N. WITTENBERG.

681.84.081.48:621.395.623

*The frequency response is all-important in gramophone reproduction, as in all fields of sound reproduction. It is not possible however, to consider the frequency characteristic of a pick-up as such, because the frequency response depends not only on the pick-up but also on the properties of the gramophone record. A further complication is that, for high frequencies, the response is dependant on the position of the pick-up on the record. These and other questions are discussed here with reference to the magnetodynamic pick-up described in Part I of the article.*

The new pick-up working on the magnetodynamic principle, type AG 3020/21 (fixed coil, moving magnet), has properties that make it well-suited to both professional and non-professional use. Its design was the subject of Part I of this article<sup>1)</sup>. In the present part of the article we shall go more closely into some properties of the pick-up.

### The frequency characteristic

The frequency characteristic of the system pick-up-gramophone record is defined, for this type of pick-up (signal proportional to needle velocity), as the signal voltage induced (r.m.s. value  $E$ ) as a function of the frequency  $f$ , when the sinusoidal recording of the test record is such that its generating point has constant peak velocity  $\hat{v}$ . From this characteristic we may deduce the required frequency response of the associated amplifier in order that the loudspeaker shall give as faithful a reproduction as possible of the original recorded sound.

Most record manufacturers work with the characteristic marked  $B$  in fig. 1 in the making of long-playing records. The curve gives the relationship between frequency and the recording velocity to be recorded on the disc (i.e. constant input voltage on the first amplifier). Since the voltage from a magnetodynamic pick-up is proportional to needle velocity, for a record cut in this manner the characteristic of the reproducing channel must be the reciprocal of  $B$ ; this is the curve marked  $A$  in fig. 1<sup>2)</sup>.

Using a test record<sup>3)</sup> with stepwise frequency

bands and recording velocities in accordance with  $B$ <sup>4)</sup> — it is possible to determine the frequency characteristic of a pick-up in combination with this record by measuring the voltage produced at each of the recorded frequencies, which are known.

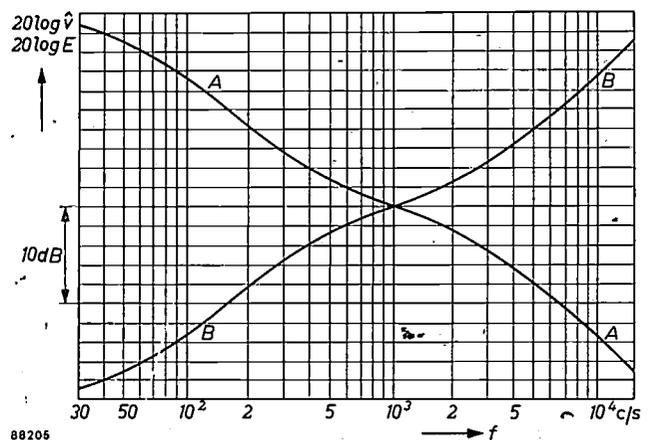


Fig. 1.  $A$ : frequency characteristic of a reproduction channel for gramophone records, largely accepted as standard.  $B$ : the reciprocal of  $A$ ; this curve shows how the modulation velocity of gramophone records should vary with  $f$ , the recorded frequency.

### Reasons why the response should extend to very high frequencies

In designing the magnetodynamic pick-up the aim was to extend the frequency response to the highest possible frequencies. One may well ask what purpose is served in trying to reproduce frequencies over 15 kc/s, since few persons apart from children can hear these frequencies, and in any case they occur only very rarely in the average recording. The reasons are the following.

In an earlier article<sup>5)</sup> in this Review it was

<sup>1)</sup> N. Wittenberg, A magnetodynamic gramophone pick-up, I. Construction, Philips tech. Rev. 18, 113-122, 1956/57.

<sup>2)</sup>  $A$  is the proposed "standard reproduction channel"; see Philips techn. Rev. 17, 104, 1955/56, or G. Slot, "Microphone to ear", Philips technical Library, Popular Series, 1956, pp. 46-49.

<sup>3)</sup> Test records of various types are available; the sound is recorded with a known modulation velocity, and the pitch may change continuously or in steps.

<sup>4)</sup> The Decca test record LXT 2695 meets this requirement satisfactorily.

<sup>5)</sup> J. B. S. M. Kerstens, Mechanical phenomena in gramophone pick-ups at high audio frequencies, Philips tech. Rev. 18, 89-97, 1956/57 (No. 3)

demonstrated that the frequency spectrum of the output voltage of a pick-up has an upper limit fixed by one of two mechanical phenomena. One of them is connected with the fact that the needle point causes elastic deformation of both groove walls (also some plastic deformation, but this is neglected here), the deformation being deeper in the convex wall (for frequencies below the needle free resonant frequency). One consequence of this is that the amplitude of the needle point is less than the amplitude of the groove, i.e. there is a *static tracing loss*. The loss is greater the shorter the wavelength  $\lambda$  recorded on the disc; at a certain wavelength  $\lambda_{co}$ , the cut-off wavelength, the loss is so great that the output voltage drops to zero. The other phenomenon that may impose an upper limit on the frequency spectrum is *resonance*. This can be either a resonance of the pick-up assembly itself<sup>6)</sup> or it may be the groove/needle resonance (stiffness of the groove wall/effective mass at needle point) dealt with in the article cited in footnote<sup>5)</sup>.

In many pick-up designs the frequency spectrum is limited by one of the resonances of the pick-up assembly, but in the magnetodynamic pick-up these appear only at frequencies higher than that of groove/needle resonance. It is therefore the groove/needle resonance that imposes the upper limit on the spectrum of the magnetodynamic pick-up. In the article<sup>5)</sup> its effect is referred to as *dynamic tracing loss*.

Now, to answer the question as to the purpose of reproducing frequencies above 15 kc/s, let us assume that it is a resonance frequency  $f_{res}$  that forms the upper limit to the spectrum, which therefore does not go up to frequencies where static tracing loss is dominant. If  $f_{res}$  falls within the audible range the result is that reproduction is "coloured", that is, a sound with the frequency of the resonance is heard continuously in the background, the sound being stronger the less the damping. This "colouring" is due to sounds at this particular frequency being emphasized by resonance. It is equally due to ringing effects each time the resonant system is excited. Not only music and speech are subject to "colouring" but also the *noise*, however faint, that is always present to some extent in the record. Accordingly, the noise heard in reproduction will be less *troublesome*, if not less strong, *the higher the limiting resonance frequency is made*.

A second argument for raising  $f_{res}$  above the audible range is that large phase shifts occur in the

neighbourhood of that frequency. These affect the reproduction of the sudden bursts of sound that frequently occur in music; phase shifts cause these transients to be lost to a great extent in reproduction.

As the spectrum is broadened out, the easier it becomes to pick out individual instruments. To some extent the effect is analogous to the better and better impression of reality given by looking at a series of half-tone photographs of the same scene but printed from blocks of increasingly finer screens. This is a third reason for putting  $f_{res}$  above the audible range.

But even if the lowest resonant frequency lies at, say, 20 kc/s, there is still no guarantee that the needle will faithfully follow the groove at very high audio frequencies, should these be present in the recording. Static loss has the consequence that, with increasing frequency, reproduction gradually falls off; at  $f_{co}$  and beyond, it inevitably drops to zero,  $f_{co}$  being the frequency corresponding to the cut-off wavelength  $\lambda_{co}$ . In general, the frequency  $f$  and the wavelength  $\lambda$  are connected by the formula  $f = V_g/\lambda$ .  $V_g$  is the velocity with which the groove passes under the needle, and it is proportional to the diameter of the circle formed by the groove being traced. Hence, as the pick-up moves inwards towards the centre of the disc,  $f_{co}$  gradually becomes lower. If  $f_{co}$  falls below the resonant frequency that has previously imposed an upper limit to the spectrum, from that moment onwards it is  $f_{co}$  that imposes the limit.

We shall now investigate these phenomena as they affect our pick-up, in the light of measurements carried out upon it.

#### *Measurements on the magnetodynamic pick-up at high audio frequencies*

Fig. 2 gives the frequency characteristic of the AG 3020 pick-up as obtained by measurement with the aid of two test records<sup>7)</sup>; on these records frequency bands of 20 kc/s, 18 kc/s, 16 kc/s, etc. are recorded with wavelengths such that there is no noticeable tracing loss, and with constant recording velocity (except for the 20 kc/s band where, according to the manufacturer, the recording velocity was 11 dB too low; it was indeed found that at 20 kc/s the voltage given was about 11 dB too low). It will be seen from the diagram that no limiting resonance occurs below 20 kc/s; that groove/needle resonance does not occur below that frequency is to be

<sup>6)</sup> See under the heading "The needle arm and the needle" in Part I of this article<sup>1)</sup> (p. 117), where lateral resonance and torsional resonance are discussed.

<sup>7)</sup> Deutsche Grammophon-Gesellschaft Nos. 68421 and 68439. These are 12" shellac records playing at 78 r.p.m., which together cover the range of frequencies from 20 kc/s down to 30 c/s.

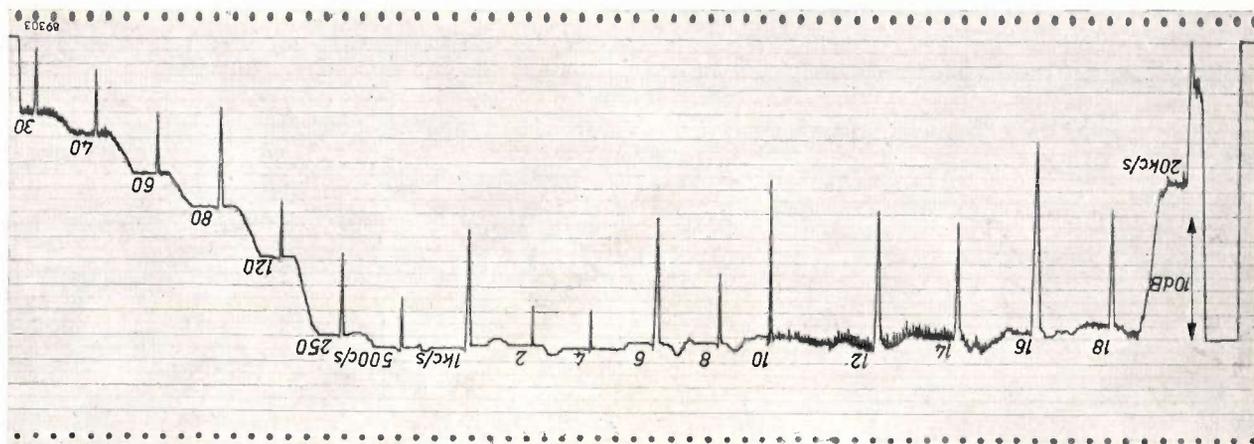


Fig. 2. Trace (on logarithmic scale) of voltage delivered at different audio frequencies by a non-loaded AG 3020 pick-up during the playing of shellac test records DGG 68 421 and DGG 68 439<sup>7)</sup>. The wavelengths cut in these discs are everywhere greater than the cut-off wavelength. The vertical force  $F_v$  was 0.1 newton. On the extreme left is the 1000 c/s voltage reference level. The measurements proper begin at  $f = 20$  kc/s (where, according to the manufacturer of the records, the modulation velocity during recording was 11 dB too low). The inverse peaks in the trace are frequency markers; only to the immediate right of each mark have the frequency and modulation velocity the right values. Modulation velocity is constant from 18 kc/s to 250 c/s and decreases at lower frequencies.

attributed, for one thing, to the high elastic modulus of the shellac material of which the test records are made.

The microgroove pick-up, AG 3021, — the same pick-up system apart from the needle — can be examined using another test record; this too has frequencies of up to 20 kc/s recorded, but with shorter wavelengths<sup>8)</sup>. The curve *a* in fig. 3 is that supplied by the manufacturer of the record and gives the recording velocity  $\hat{v}$  as a function of frequency, as determined by optical methods; also indicated in the diagram is the wavelength  $\lambda$  for each of the seven bands from 20 to 8 kc/s. An ideal pick-up would have an output voltage curve with the same shape as curve *a*. Instead, we find the characteristic marked *b*; it seems that about 18 kc/s a small resonance occurs which causes the characteristic to rise by a few decibels, after which it falls rapidly. It is not permissible to attribute the difference between *b* and *a* entirely to resonance, however, for a certain amount of static tracing loss may be involved. To separate this latter from resonance effects (dynamic tracing loss) we make use of the following artifice.

Referring to the wavelength scale (top right, fig. 3), *b* gives the voltage as a function of  $\lambda$ . In order to eliminate resonance phenomena we shall associate the same wavelengths with frequencies at which no resonance occurs by making the disc turning more

slowly. If we make the speed 10 r.p.m. (instead of  $33\frac{1}{3}$  r.p.m., the speed at which curve *b* was obtained), the highest frequency reproduced becomes 6 kc/s; the form of *b* makes it unlikely that any resonance will occur below that frequency.

The voltages measured at 10 r.p.m. are given in curve *c* as a function of wavelength (not of frequency); the ordinates have been corrected for the lower voltages measured resulting from the lower speed. Hence *c* gives, as a function of  $\lambda$ , the *needle*

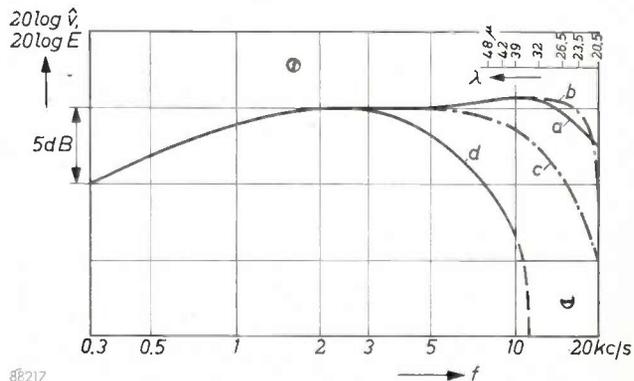


Fig. 3. *a*) Modulation velocity  $\hat{v}$  as function of frequency  $f$  for the Cook 10A synthetic resin test record (curve supplied by the manufacturer<sup>8)</sup>). *b*) Voltage  $E$  delivered by non-loaded AG 3021 pick-up during playing of Cook 10A record at  $33\frac{1}{3}$  r.p.m.; wavelength scale appears at top right. *c*)  $E$  as function of  $\lambda$ , for the same record turning at 10 r.p.m.; ordinates have been corrected to allow for the low rate of rotation. The range of wavelengths from 20.5 to 48 microns is recorded in grooves of diameters from 24 to 22 cm. *d*) curve of  $E$  as function of  $f$ , as reconstructed for the innermost groove, which has a diameter of 12.5 cm. The vertical force used in obtaining curves *b*, *c* and *d* was 0.1 newton.

<sup>8)</sup> Test record Cook 10A (made by Cook Laboratories), a 10'' synthetic resin microgroove disc playing at  $33\frac{1}{3}$  r.p.m.

point velocities occurring at the recording velocities of  $a$  but without interfering resonances. The difference between  $a$  and  $c$  is therefore the static tracing loss. In the particular conditions in which the measurements were carried out (radius of curvature of the diamond needle point  $25 \mu$ ; vertical needle force  $F_v = 0.1 \text{ N}$ ; record of synthetic resin) static tracing loss becomes perceptible at wavelengths shorter than about  $50 \mu$ .

The full curve in fig. 4 is the static tracing loss plotted as function of  $\lambda$ . The function is expressed by equation (10) of article 5):

$$\hat{y}_s = \hat{y}_{co} \left\{ 1 - \left( \frac{\lambda_{co}}{\lambda} \right)^2 \right\}$$

where  $\hat{y}_s$  is the amplitude of the needle point in the absence of resonance, and  $y$  and  $\lambda$  are the amplitude and wavelength of the groove cut in the record. By inserting various values of  $\lambda$  and the corresponding values of  $\hat{y}_s/\hat{y}$  given in fig. 4 we can determine  $\lambda_{co}$ , the wavelength at which the velocity of the needle point is zero and at which therefore the pick-up ceases to deliver a voltage. In this way we find that  $\lambda_{co} = 16 \mu$ .

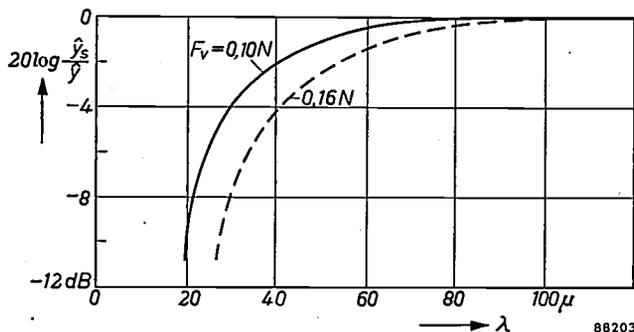


Fig. 4. The static tracing loss  $20 \log \hat{y}_s/\hat{y}$ , as derived from curves  $a$  and  $c$  in fig. 3, plotted against the groove wavelength  $\lambda$  for vertical forces  $F_v$  of  $0.10$  and  $0.16 \text{ N}$ . The cut-off wavelength becomes longer as  $F_v$  is increased.

For a given set of conditions (radius of curvature of needle point  $25 \mu$ ;  $F_v = 0.1 \text{ N}$ ; record of synthetic resin having an elastic modulus  $3.3 \times 10^9 \text{ N/m}^2$ ), equation (11) of article 5) gives the following relation between cut-off frequency  $f_{co}$  and groove velocity  $V_g$  in m/sec:

$$f_{co} = 55\,000 V_g \text{ c/s.}$$

Hence  $\lambda_{co} = V_g/f_{co} = (1/55\,000) \text{ metres} = 18 \text{ microns}$ . Considering the errors involved the measurements described, the fact that extrapolation has to be used, and the uncertainty in the value of the modulus of elasticity, which is considerably dependent on temperature, the agreement between calculation ( $18 \mu$ ) and measurement ( $16 \mu$ ) is satisfactory.

Equation (17) of article 5) states that  $\lambda_{co}$  is proportional to the cube root of the vertical needle force  $F_v$ . We can verify this relation by repeating the measurements with another value of  $F_v$ , say  $0.16 \text{ N}$ . The cut-off wavelength to be expected would then be  $(0.16/0.10)^{1/3}$  times  $16$  or  $18 \mu$ ; that is, say  $20 \mu$ . The broken-line curve in fig. 4 is drawn for an  $F_v$  of  $0.16 \text{ N}$  and from this we find a  $\lambda_{co}$  of  $22 \mu$ . Here too, therefore, there is satisfactory agreement.

Let us return to fig. 3 and compare curves  $b$  and  $c$ . The difference between them is the dynamic tracing loss, which arises on account of a resonance phenomenon at about  $18 \text{ kc/s}$ , where the difference is at its greatest. The assumption that this resonance is that of the groove/needle system (stiffness of groove wall and effective mass at needle point) can be justified as follows.

The resonant frequency of a damped vibration is given by:

$$f_{res} = f_0 \sqrt{1 - \frac{1}{2Q^2}} \dots \dots \dots (1)$$

in which  $f_0$  is the natural frequency of the system and  $Q$  is the quality factor. A good approximation for  $f_0$  is  $\sqrt{S/m}/2\pi$ . In accordance with equation (15) of article 5), the stiffness of the groove wall is:

$$S = \frac{3 F_v}{2 a_0 \sqrt{2}}$$

Taking  $F_v = 0.1 \text{ N}$  and  $a_0$  (depth of penetration in the unmodulated groove wall)  $= 2 \times 10^{-6} \text{ metre}$  we find that  $S = 53\,000 \text{ newtons/metre}$ , and taking  $m = 3 \times 10^{-6} \text{ kg}$  we obtain  $f_0 = 21.2 \text{ kc/s}$ . From fig. 3 we see that the greatest difference between  $b$  and  $c$  is  $6 \text{ dB}$ , which corresponds to a  $Q$  of  $2$ . These values of  $f_0$  and  $Q$ , inserted in equation (1), produce a  $f_{res}$  value of  $19.8 \text{ kc/s}$ , which is near enough to the resonance observed at  $18 \text{ kc/s}$  to make it probable that this resonance is that of the groove/needle.

We see from fig. 3 that the resonance effect is stronger than might be mistakenly supposed from curves  $a$  and  $b$ , and that over a big range of frequency it compensates the static tracing loss. The question might be posed as to whether the straightness of the characteristic in fig. 2 is also due to compensation of this kind. This does not seem to be the case: if we give a slower turning speed to the shellac record with which this characteristic was obtained, we get the same line as that given at the normal speed of  $78 \text{ r.p.m}$ . The shortest wavelength on this record is  $60 \mu$ , this producing the  $20 \text{ kc/s}$  tone; this frequency is cut in a groove of about  $30 \text{ cm}$  diameter. Fig. 4, which was obtained with a synthetic resin record, shows that the tracing of this

wavelength involves a loss smaller than 1 dB, so that for the shellac record, with its higher elastic modulus, it is certain that no perceptible tracing loss will arise at  $\lambda = 60 \mu$ . We may therefore conclude from the straightness of the characteristic that the system is free of resonance below 20 kc/s.

Comparison of curves *a* and *b* in fig. 3 may perhaps lead to the question: what is the significance in practice of the static tracing loss? The loss is, after all, largely compensated by dynamic tracing loss (groove/needle resonance), which is determined by the material of the record, the radius of curvature of the needle, the vertical force  $F_v$  and the effective mass — that is to say by factors that are always the same for a given gramophone and type of record. Apart from this, it would be an easy matter to correct insufficient or excessive compensation in the amplifier. But what has to be borne in mind is that static tracing loss is a *direct* function of wavelength, and only an *indirect* function of frequency, via the relation  $\lambda = V_g/f$ . Consequently the frequency at which a given combination of record and pick-up will exhibit a certain static loss is dependent on the groove velocity  $V_g$  and thus on the diameter of the groove.

On the record used for the measurements of fig. 3, the range of frequencies from 20 kc/s to 10 kc/s is registered in grooves having a diameter of from 24 cm to 22 cm; consequently, curve *b* in fig. 3 is valid only for that portion of the disc. Further towards the edge of the record the same wavelengths produce higher frequencies; here, therefore, there is stronger reproduction of high frequencies than from grooves nearer the centre. It would be possible to obtain the characteristic for the innermost grooves by using a test record on which wavelengths of from 20 to 50  $\mu$  were cut in grooves having a small diameter. Such a disc is not available, but we can reconstruct the characteristic in question from curves *a* and *c* by taking into account that the frequency at which static tracing loss has a given value is proportional to the diameter of the groove. Characteristic *d* in fig. 3 has been constructed in this manner; it applies to the innermost groove of a long-playing record, which has a diameter of 12.5 cm. The cut-off frequency, which is 23-26 kc/s for curve *b* ( $\lambda = 18$  or  $16 \mu$ ), is only 12-13.5 kc/s for curve *d*. These frequencies are below the resonant frequency and hence there can be no compensation due to resonance (negative dynamic tracing loss) above 12-13.5 kc/s in the innermost grooves.

The foregoing has attempted to show that there is no point in supplying a frequency characteristic with a pick-up without at the same time specifying the

test record used and the value of  $F_v$  the vertical force during the measurements. Even then, the characteristic gives no more than an overall impression of what may be expected from the playing of a normal record; it must be remembered that static tracing loss is dependent on the groove velocity, that is, on the diameter of the groove being traced. Fig. 3 shows that this effect becomes perceptible at a frequency as low as 3 kc/s.

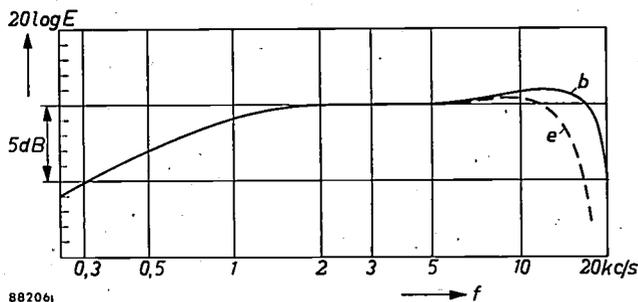
#### *Reproduction of very low frequencies*

In the foregoing attention has only been given to the upper register. We shall now say a word about the reproduction of very low frequencies. In the magnetodynamic pick-up, unlike the electrodynamic there is no objection to the use of a coil with a large number of turns. An input transformer — essential for the electrodynamic pick-up — is thereby rendered superfluous. The transformer is an expensive item if good reproduction of low audio frequencies is required for its primary then has to have a high self-inductance. The fact that there is nothing to set a lower limit to the frequency spectrum of the magnetodynamic pick-up is, of course, favourable for the reproduction of musical sounds of the lowest pitch, but at the same time it leaves the way open for undesired low frequencies. Vibrations of very low frequency arise through mechanical shock or vibration; they may arise in the gramophone itself or in the recording equipment (in the latter case they form part of the recording). Although these frequencies are often below the limit of audibility (in the region of 10 c/s, say), they can be very troublesome in that the unwanted voltage they cause may take up a large part of the grid swing in the output stage; when this happens even a weak signal, in combination with the unwanted voltage, may overload the amplifier. The difficulty is removed by suitable design of the amplifier, but to go into details would be to depart from the framework of the present article.

#### *Effect of the electrical load on the frequency characteristic of the pick-up*

The load across the pick-up during the above determinations of the characteristics was very small. In normal use, however, there is a greater load on the pick-up — not only that of the amplifier input impedance, but also that of a capacitance of about 250 pF, consisting mainly of the capacitance of the connecting lead between pick-up and amplifier. Since the internal impedance of the pick-up is partly due to a self-inductance of 0.6 H, a tuned circuit is formed; the resonant frequency of

which is in the vicinity of 13 kc/s; this resonance, being inside the audible range, is undesirable. By making the right choice of load impedance, however, this resonance can be damped to a point where it practically ceases to be effective. Two frequency response curves appear in *fig. 5*: the first is curve *b*



*Fig. 5.* The effect of the pick-up's electrical load on its frequency characteristic. Curve *b* was obtained with a pick-up type AG 3021 subject to practically no load; for curve *e* the pick-up had a load consisting of 68 k $\Omega$  in parallel with 250 pF, the circumstances otherwise being the same. Curve *b* here is identical to curve *b* in *fig. 3*.

from *fig. 3*, as obtained with a parallel capacitance of some tens of pico-farads and a load impedance of 0.5 M $\Omega$ ; curve *e*, the second characteristic, was obtained by using the same test record with a parallel capacitance of 250 pF and a load impedance of 68 k $\Omega$ . The difference between the two curves becomes greater than 2 dB only above 14 kc/s. With the majority of records the difference is scarcely audible; for this reason a value of 68 k $\Omega$  is recommended for the terminating resistance.

#### Non-linear effects

When short wavelengths — that is, high frequencies cut in small-diameter grooves — are traced, effects become perceptible which, for the sake of

simplicity, were not taken into consideration in the article referred to under <sup>5)</sup>. Thus strong non-linear tracing distortion may arise <sup>9)</sup> because, for example, the peaks of a sinusoidal groove are so sharp that the needle point no longer fits into them. The main consequence is that third harmonics are set up. Even if their frequencies are above the cut-off frequency, it is still possible for these harmonics to be reproduced, for the cut-off frequency is determined by the static tracing loss which itself depends only on the actual pattern cut in the disc: hence any frequencies due to distortion in the playback process are not effected by the cut-off frequency. This non-linear effect may partially compensate harmonics recorded on the record but lost on account of static tracing loss, and in this way it may give rise to an impression of better reproduction of the upper register. However, these phenomena have been so little investigated that it is not possible to go into them here.

<sup>9)</sup> J. A. Pierce and F. V. Hunt, On distortion of sound reproduction from phonograph records, *J. Acoust. Soc. America* **10**, 14-28, 1938/39. W. D. Lewis and F. V. Hunt, A theory of tracing in sound reproduction from phonograph records, *ibid.* **12**, 348-365, 1940/41.

**Summary.** Part I of this article dealt with the design of the magnetodynamic pick-up, types AG 3020/21. In Part II above the author puts forward a number of arguments for extending the frequency response to above 15 kc/s, as has been done in the magnetodynamic pick-up. In the course of the discussion on the frequency characteristic for the upper register, a method is described of separating static and dynamic tracing losses, these being a wavelength and a frequency effect respectively. With records of synthetic resin, static tracing loss exhibits a cut-off frequency of about 24 kc/s where the groove has a diameter of 24  $\mu$ m, and of about 12 kc/s for the innermost groove; dynamic tracing loss, on the other hand, is connected with a groove/needle resonance at about 18 kc/s. A word is also devoted to reproduction at very low frequencies, to the effect of the electrical load on the pick-up and to non-linear effects.

## ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk \* can be obtained free of charge upon application to the Philips Research laboratory, Eindhoven, Netherlands.

2322: G. W. Rathenau and J. F. Fast: Initial permeabilities of sintered ferrites (*Physica* **21**, 964-970, 1955).

It has been shown (Wijn, Went) that the initial permeability of sintered NiZn ferrites is mainly caused by simultaneous rotation of the spins. This view is supported by the present investigation, especially by the evidence of experiments in which the spin distribution was anisotropic in the absence

of external stresses. Also, the influence of stress on the initial permeability can be accounted for by considering simultaneous spin rotation only. Under certain circumstances a contribution of domain wall movement to the initial permeability must be taken into account. This is the case, for example, when the material has been demagnetized by heating to temperatures above the Curie temperature and cooling.

**2323:** A. H. Boerdijk: Het mechanisch evenwicht van een lichaam dat over een vlak kan rollen (Simon Stevin 30, 193-231, 1955). (The equilibrium of a body that can roll on a surface; in Dutch).

A rigid body  $B'$  is placed upon a rigid surface  $O$  in a position of mechanical equilibrium with respect to a gravitational force acting on the body. The stability of this equilibrium is investigated for the case of small rolling or spinning movements of  $B'$  over  $O$ . Sufficient conditions for stability and lability are expressed inter alia in terms of the principal radii of curvature of  $O$  and of the surface  $B$  of  $B'$  at the contact point or points. It is shown that there exists an infinite number of combinations of bodies and surfaces showing neutral equilibrium, the case of a sphere on a horizontal plane being a special one. If either  $B$  or  $O$  are given, neutral equilibrium is always possible if there is more than one point of contact. In the case of one point of contact neutral equilibrium is possible only if  $B$  and  $O$  satisfy certain partial differential equations.

**2324:** H. J. J. van Boort and M. Klerk: High frequency oscillations in low pressure mercury arcs (Appl. sci. Res. B5, 69-74, 1955).

The influence of the associated circuit of a low-pressure tubular fluorescent lamp on the propagation of the noise generated in the lamp is discussed. Two types of noise, cathode noise and anode noise, have been distinguished by the use of an electronic method. The cathode-noise is subjected to a more detailed investigation, particularly the re-ignition noise. A hypothesis is put forward regarding the manner of generation of two different types of re-ignition noise.

**2325:** W. Elenbaas: The behaviour of the high pressure discharge in Xe-Hg mixtures (Appl. sci. Res. B5, 205-209, 1955).

Spectral data and voltage gradients have been determined for a series of Xe-Hg discharges at constant loading. With increasing Hg%, voltage gradient and Hg spectrum intensity increase whilst the Xe spectrum intensity decreases. The results are explained by a consideration of the variation in discharge temperature, leading to theoretical values in good agreement with experimental results.

**2326:** H. G. Beljers: Application de la résonance ferromagnétique dans les ferrites (Cahier de Phys. No. 62, 43-44, October 1955).

With magnetically polarized ferromagnetic materials such as ferroxcube, the principle of reci-

procity may not be obeyed. Transmission of electromagnetic waves through such a material gives rise to the Faraday effect or the Cotton-Mouton effect, depending on the mutual orientation of the transmission and the polarization. Practical applications of both possibilities are briefly indicated.

**2327:** J. S. van Wieringen: La résonance ferromagnétique des ferrites à température de compensation (Cah. de Physique 45-47, 1955).

Continuation of the work reported in abstract 2114. The gyromagnetic factors of Li-Fe-Cr spinels of two different compositions are given at 9500 and 24000 Mc/s.

**2328:** W. J. Oosterkamp, G. M. Ardran and others: Discussion on "image intensification in radiology" (Proc. Instn. Electr. Engrs. 102B, 845-849, 1955).

Short discussion of design factors, quantum fluctuations and noise, problems in medical applications, etc. of X-ray image intensifiers.

**2329:** C. Z. van Doorn and Y. Haven: Dichroism of the  $F$  and  $M$  absorption bands in KCl (Physical Review 100, 753, 1955).

An additively coloured KCl crystal showing the  $F$  and  $M$  absorption bands becomes dichroic after irradiation in the  $F$  band with polarized light at 77°K. Both the  $F$  and  $M$  bands show this dichroism. A crystal which shows only the  $F$  absorption band does not become dichroic under the same circumstances. These and other experiments indicate that the  $M$  centres responsible for the  $M$  absorption are anisotropic and are oriented along the 2-fold symmetry axes. In crystals showing both  $F$  and  $M$  absorption bands, at least part of the  $F$  band must be due to  $M$  centres instead of  $F$  centres. The dichroism is attributed to reorientation of the  $M$  centres by ionic movement induced by irradiation in the  $F$  band.

**R 292:** S. Landsberg: A general purpose drift-free D.C. amplifier (Philips Res. Rep. II, 161-171, 1956).

A method for drift reduction in D.C. amplifiers is described. The advantage of the method is that the circuit which stabilizes the output voltage of the D.C. amplifier does not interfere with its characteristics. This is accomplished by comparing a fraction of the output voltage with the input voltage by means of a mechanical chopper. The chopped difference is amplified by an A.C. amplifier and its rectified output voltage is fed to the D.C. amplifier,

so as to reduce the drift. Thus, the zero-point control circuit acts as a voltage stabilizer, with the input serving as a reference voltage. An experimental amplifier constructed on this principle has a gain of 8000, flat from D.C. to 300 kc/s. Input impedance at D.C. is of the order of 100 M $\Omega$ . The stability is of the order of 10  $\mu$ V (with respect to the input terminals) over a period of more than five hours.

**R 293:** K. F. Niessen: Spontaneous magnetization and magnetic susceptibilities of an antiferromagnetic with foreign ions in only one sublattice (Philips Res. Rep. **11**, 172-182, 1956).

A study is made of an antiferromagnetic in one of whose sublattices a relatively small number of original *A*-ions is supposed to be replaced by ions of another metal *B*, the moment *m'* of *B* being different from the moment *m* of *A*. The magnetization is then no longer zero. Its dependence on the temperature is investigated and also the change in the parallel and perpendicular susceptibilities due to the above replacement.

**R 294:** O. W. Memelink: The distribution of impurity in a semi-infinite solidified melt (Philips Res. Rep. **11**, 183-189, 1956).

The transient distribution of impurity in a solidifying melt is calculated for the case of a homogeneous initial impurity concentration and constant velocity of the solid-liquid interface. The assumptions involved are: (1) the fraction of impurity concentration segregating from the liquid at the interface is constant, (2) the transport of solute from and to the interface occurs by diffusion on the liquid of semi-infinite length. The transient distribution for various values of the segregation constant is evaluated numerically.

**R 295:** F. K. Lotgering: On the ferrimagnetism of some sulphides and oxides (Philips Res. Rep. **11**, 190-249, 1956).

The ferrimagnetic behaviour of a number of sulphides with nickel arsenide and with spinel structure has been studied. Iron sulphides of compositions between FeS and Fe<sub>7</sub>S<sub>8</sub> (all nickel arsenide structure) show antiferrimagnetic properties

with low sulphur content and ferrimagnetic properties with high sulphur content. These are investigated experimentally and explained in terms of order-disorder phenomena. Magnetic measurements were made on MnCr<sub>2</sub>S<sub>4</sub>, FeCr<sub>2</sub>S<sub>4</sub>, CoCr<sub>2</sub>S<sub>4</sub> and ZnCr<sub>2</sub>S<sub>4</sub> (all spinel structure). The properties of these sulphides and those of the corresponding oxides are discussed in relation to Yafet and Kittel's theory of ferrimagnetism. The theory is extended to cover normal spinels AB<sub>2</sub>X<sub>4</sub> with one type of A-ion and one (different from A) type of B-ion.

**R 296:** J. Bloem: Controlled conductivity in PbS single crystals (Philips Res. Rep. **11**, 273-336, 1956).

Lead sulphide single crystals were reheated under various sulphur pressures in order to influence the composition of the crystals. Electric and thermoelectric measurements after quenching showed that PbS crystals are *n*-type (conduction by electrons, surplus Pb) after heating under low sulphur pressures and *p*-type (conduction by holes, surplus S) after heating under high sulphur pressures.

For crystals doped with small concentrations of Bi or Ag, there appears to be a range of atmospheres, where the electron or hole concentrations are equal to the concentration of added Bi or Ag ions (range of controlled valency); this range is broader the lower the heating temperature. The experimental results could be explained quantitatively using the theory of Schottky and Wagner on lattice imperfections as applied by Kröger, Vink and Van den Boomgaard. In this way it has been possible to achieve some understanding of the properties of this compound; the theory may well, in principle, also be applicable to other semiconducting compounds.

**R 297:** F. K. Lotgering: On the ferrimagnetism of some sulphides and oxides (Philips Res. Rep. **11**, 337-350, 1956).

Continuation of **R 295**. Magnetic properties of MnCo<sub>2</sub>O<sub>4</sub>, NiCo<sub>2</sub>O<sub>4</sub>, Co<sub>3</sub>O<sub>4</sub>, ZnCo<sub>2</sub>O<sub>4</sub>, Co<sub>3</sub>S<sub>4</sub> and NiCo<sub>2</sub>S<sub>4</sub> have been measured and are discussed in connection with the character of their chemical bonds.

# Philips Technical Review

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

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

## THE LIFE AND RELIABILITY OF VALVES

by K. RODENHUIS, H. SANTING and H. J. M. van TOL.

621.385:621.375.2

*In recent years there has been a substantial increase in the use of electronic equipment. Not only have telecommunications developed enormously, but new fields of application have been opened up. Typical examples are industrial measurement and control equipment, electronic computers and many new kinds of navigational aids for ships and aircraft. A modern airliner, with radar, may have as many as 500 valves in its equipment. Moreover, the military use of electronic equipment has expanded on a vast scale. American publications mention figures like 2000 electronic valves in a large bomber and 10 000 on board a battle-ship.*

*As a result of the rapidly increasing use of electronic valves in industrial equipment, aircraft and etc., far more attention is being paid to the reliability of these components than ever before, particularly in view of the often serious consequences of a valve failure in complex equipment. In the following article new definitions are given for the related concepts "reliability" and "life". A description follows of the measures adopted in the factory, and of those which the user can adopt, to improve the reliability and life of valves.*

Originally, the electronic valve was used predominantly in the field of broadcasting, later extended to include television. While both these applications are still expanding, there has been in recent years an enormous increase in the use of valves in other fields, here referred to quite generally as "professional" applications. These may be classified under four main headings: carrier-telephony, electronic computers, industrial measurement and control equipment, and mobile transmitters and receivers. Several modern "Special Quality" valves for these applications are shown in *fig. 1*.

Owing to the serious consequences that may arise from the failure of a valve in professional equipment, the conventional requirements — for example with regard to the mutual conductance, power dissipation, suitability for wide-band amplification, etc. are no longer adequate. It is evident that valves used for professional equipment should satisfy the additional requirement of reliability, for the equipment should be ready for operation at all times and there must be no risk of sudden failures. Other properties that may be important are long life, shock resistance and narrow tolerances in electrical characteristics, although they may not all be equally necessary in every application. For instance, long life and constant electrical characteristics are required of valves used in computers and in telephone repeater amplifiers, but ruggedness

is not so necessary, for in this equipment the valves are not subjected to mechanical vibrations or shocks. On the other hand, the valves used in aircraft transmitters and receivers must be able to withstand severe vibrations and shock, but some variation in characteristics is usually to be tolerated, and life also presents no problems where the equipment is in operation only for short periods a day.

The terms "reliability" and "life"

It is necessary in the first place to define the terms "reliability" and "life". To do this, we postulate an electronic apparatus equipped with a large number  $S_0$  (e.g. 1000) of new valves, all of the same type, and which at a given moment is switched on for a long period of operation. Each valve failure is recorded, so that the number of original valves still functioning ( $S$ ) can be seen at any instant ( $t$ ). If  $S/S_0$  is plotted, on a logarithmic scale, as a function of  $t$ , a curve is obtained similar to those shown in *fig. 2*, which have all been taken from publications on the subject <sup>1)2)3)4)5)</sup>.

<sup>1)</sup> N. W. Lewis, Post Office Electr. Engr's J. 41, Part I, 10-12, 1948.

<sup>2)</sup> D. K. Gannet, Bell Labs. Record 18, 378-382, 1940.

<sup>3)</sup> C. R. Knight, A.I.E.E. Conf. Electron tubes for instrumentation and industrial use, Philadelphia, March 1948.

<sup>4)</sup> Eleanor M. MacElwee, Sylvania Technologist 3, 16-20, April 1950.

<sup>5)</sup> K. Rodenhuis and W. Sparbier, Elektron. Rundschau 9, 22-25, Jan. 1955 and 72-74, Febr. 1955.

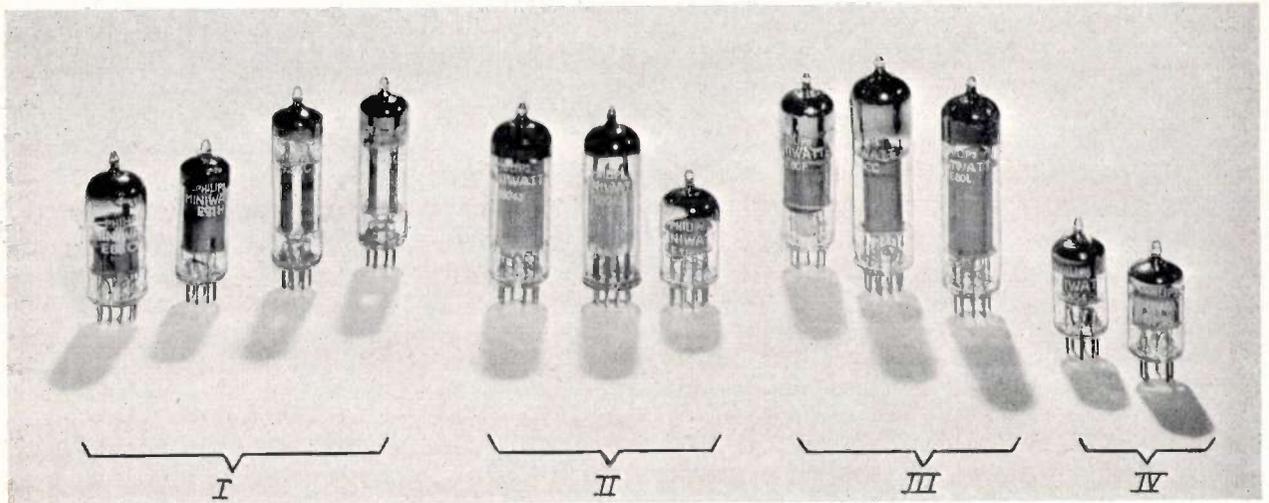


Fig. 1. Some modern "Special Quality" valves for professional equipment.  
 Group I. Valves for electronic computers (great reliability, long life (also when operated cut-off), narrow spread in characteristics affecting switching functions): E 80 CC, E 91 H, E 92CC, E 90 CC.  
 Group II. Valves for telephone repeater amplifiers (great reliability, long life, narrow tolerances in amplifying, characteristics): 18045, 18042, E 180 F.  
 Group III. Valves for control and metering equipment (same characteristics as in group II but also shock-proof): E 80 F, E 80 CC, E 80 L.  
 Group IV. Valves for mobile radio stations (great reliability, shock-proof): 5654, 5726.

88729

Curve I found by Lewis<sup>1)</sup> is noteworthy; it is almost a straight line, having the equation:

$$S/S_0 = e^{-Pt}, \dots \dots \dots (1)$$

where  $P$  is a constant. From this simple expression certain interesting conclusions may be drawn. For example, the number of valve failures per unit

time, i.e.  $-dS/dt$ , is  $PS_0 e^{-Pt} = PS$ . Consequently, the percentage failures per unit time (the "failure rate", analogous to the death rate of a population), i.e.  $-100 (dS/dt)/S$ , is equal to the constant  $100 P$  and is therefore independent of time.

In the case of curve I, the average life  $L_m$  of the group is found from the equation:

$$L_m = \frac{1}{S_0} \int_0^{\infty} S_0 e^{-Pt} dt = \frac{1}{P} \dots \dots (2)$$

For curve I,  $P = 0.135 \times 10^{-3}$  per hour (failure rate 13.5% per 1000 h) and therefore  $L_m = 1000/0.135 = 7500$  h.

At a time  $t = L_m$ , the number of survivors is:

$$S_{L_m} = S_0 e^{-1} = 0.368 S_0,$$

in other words, 36.8% of the original number are still in operation. At a time  $t = 2 L_m$ , the number of survivors is  $S_{2L_m} = S_0 e^{-2}$ , and so on: in each period  $L_m$  the percentage survivors decreases by a factor  $e$ . This means that the valves that have already been in operation for thousands of hours are in no respect to be distinguished from brand-new valves, and that there is therefore no point in replacing them. Obviously, this holds good only for a straight line, such as curve I in fig. 2.

Lewis<sup>1)</sup> has pointed out that a straight line is only to be expected where completely random conditions prevail. It cannot therefore be taken as representative of the normal case. The type of

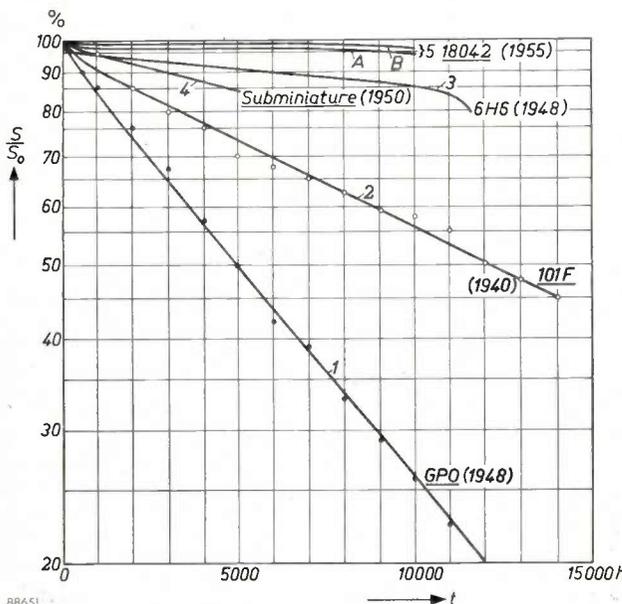


Fig. 2. The percentage  $S/S_0$ , on logarithmic scale, as a function of time  $t$ , for life tests on various types of valve. The curves 1...5 have been taken from published sources<sup>1) . . . 5)</sup>; the valve type and the year of publication of the curves are given in each case. The data for curves 5A and B were obtained from a test made with the cooperation of the Netherlands Post Office.

88651

curve that should be regarded as normal is, in our opinion, curve 3, given by Knight<sup>3</sup>). This shows a sharp drop during the first few hundred hours, followed by a long, fairly straight portion with a slight slope, which finally falls steeply again after many thousands of hours. The form of this curve may be explained as follows. In the beginning, some valves show the effects of manufacturing faults not detected in the factory, such as bad welds or near short-circuits between electrodes. After these valves have been removed, there will be merely an occasional "random" failure; until finally certain kinds of predictable defects become prevalent, defects arising from gradual physical and chemical processes in and near the cathode, such as declining emission and diminishing insulation resistance between electrodes.

A curve of this form (3, fig. 2) yields the following information.

- 1) In the straight central portion of the curve the valves behave more or less in accordance with the exponential law (eq. 1).
- 2) The higher failure rate during the first 100 to 1000 hours means that valves that have been in operation for several hundred hours are more reliable than new valves.
- 3) Just past the straight portion, where the failure rate increases, it is advisable (unless certain precautions have been taken, which will be discussed later) to replace all survivors by new valves.

As regards the latter point we shall define the end of a valve's "useful life" ( $L_p$ ) as the moment at which the failure rate begins to increase (fig. 3). This moment depends, of course, not only on the quality of the valve but also on the circuit, since the permissible deviations that may occur in cer-

tain valve characteristics are determined by the particular function of the valve in the circuit. Two important valve-characteristics in this respect are mutual conductance and grid current. To give a typical objective criterion: a valve is regarded as unserviceable when its mutual conductance has dropped to 70% of the nominal value, or when its grid current (under specific conditions) has risen to 1  $\mu\text{A}$ . Other criteria may be chosen for valves used in special equipment, for example in electronic computers.

Other definitions of "life" have been given by Eleanor MacElwee<sup>4</sup>), who regards it as the time after which 80% of the valves are still functioning, and by Gannet<sup>5</sup>), who takes the average life according to (2) as the life of a valve.

Gannet's definition is useful for exponential characteristics (straight lines in fig. 2) but for more representative curves, such as 3, it gives a result of little practical value. According to MacElwee's definition a "life of 5000 hours" represents in fact a failure rate of about 4% per 1000 h. The present authors prefer to relate the failure rate to reliability rather than to life (see below).

During a practical life test the failure rate (except at the beginning) is constant and relatively low. This failure rate mainly determines the frequency of the troubles arising from valve breakdowns. We shall regard the reciprocal value of this failure rate as a measure of *reliability*; the flatter the curve the greater the reliability.

The behaviour of a valve of great reliability is exemplified by curves 5A and 5B in fig. 2, both of which apply to a type 18042 pentode, used in telephone repeaters. The two curves refer to different applications. Curve 5A was plotted from data recorded during a test on a circuit highly sensitive to insulation defects in valves. In this case, each valve was signalled as a failure, via a relay, as soon as its insulation resistance fell to 1 M $\Omega$ . Curve 5B applies to the same type of valve under the same load conditions, but employed in an ordinary AF amplifier where a drop in insulation resistance to 1 M $\Omega$  could do no harm. In the right portion, curve 5A has a slope of 0.5% per 1000 h and curve 5B a slope of 0.25% per 1000 h. The curves still showed no tendency to bend over even at 10 000 hours.

It is believed that these figures do not represent the utmost that can be reached. It may be expected that with further experience it will be possible to reduce the failure rate of Special Quality valves to about 0.1%, a figure that should represent a reliability sufficient for practically all purposes.

A distinction may be made between the sudden and the gradual failure of a valve. A sudden failure without previous warning may be due to a broken connection, short-circuiting, or a cracked envelope.

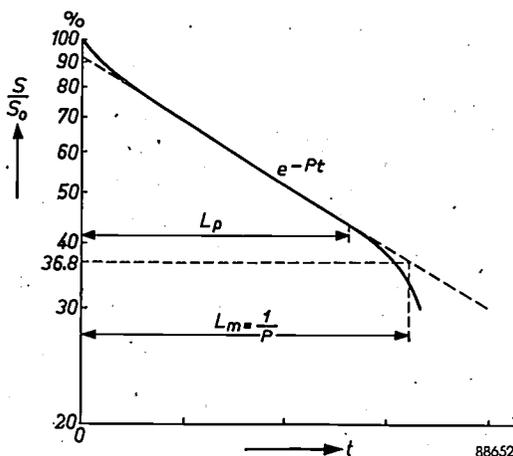


Fig. 3. Illustration of the terms "average life"  $L_m$  and "useful life"  $L_p$  ( $L_p$  comes to an end when the failure rate begins to increase).

A gradual failure implies, for example, slowly declining emission (associated with a drop in mutual conductance and in output power), deteriorating insulation, rising grid current, etc. Gradually failing valves in an equipment can be detected beforehand by regular measurements, and serious breakdowns can be avoided by replacing such valves in good time. The advantage of this "preventive maintenance" as against replacement en bloc is that the valves are replaced only as the need arises. The disadvantage — particularly where the plant concerned is equipped with large numbers of valves — is that it may entail putting the equipment partly or wholly out of operation during the measurements. Methods have been devised, however, which allow the equipment to remain in operation during inspection. For this purpose the working conditions of the valve are deliberately worsened, so that any shortcoming can be more readily detected. In a telephone repeater, for example, the test can take the form of decreasing the heater voltage or increasing the resistance in the grid circuit and then measuring the effect on the anode current; this can be done without taking the valve out of its socket or interrupting its operation. In electronic computers so-called "marginal checking" methods are used, in which it is ascertained by how much the amplitude or repetition frequency of control pulses can be varied before the computer begins to make mistakes.

Efficient preventive maintenance not only forestalls breakdowns but effects considerable economies, since it enables the valves to be used up to the end of their individual lives.

Both types of failures, gradual and sudden, will now be dealt with in turn. A discussion then follows of the contacts between valve and valve holder, electrical tolerances, ruggedness and quality control.

### Gradual failures

#### *Cathode phenomena*

Most valves have an indirectly heated oxide cathode, which consists of a nickel tube coated with a mixture of crystals of BaO and SrO (and sometimes CaO). A tungsten wire inside the nickel tube heats the cathode to a temperature of 750 to 800 °C<sup>6</sup>). The saturation emission of such cathodes lies between 2.5 and 25 A/cm<sup>2</sup>.

A high emission is dependent on the oxide mixture having a low work function, and this is the case only when a small fraction of Ba atoms are present in the oxide mixture (of the order of 0.01%) not

combined with oxygen. Free Ba atoms may originate in two ways:

- a) Reaction of the BaO with reducing agents in the nickel. These are minute impurities in the cathode nickel that constitute less than 1% of the total weight but affect the emission properties to a very large extent.
- b) Electrolysis of the oxide layer by the cathode current passing through it.

During operation, however, certain processes take place in the valve that tend to decrease the quantity of free barium. If the free barium is removed faster than it is produced the result will be a gradual decline in emission. The processes are of two kinds, physical (evaporation) and chemical ("poisoning").

Poisoning may be due to residual gas. Even with efficient pumping and getter action, a residual pressure of about 10<sup>-7</sup> mm of mercury remains in the valve. The number of gas molecules in the valve is then 10<sup>-10</sup> less than at atmospheric pressure, but there are still 3 × 10<sup>9</sup> per cm<sup>3</sup>. Moreover, gas may also be given off gradually by the electrodes and the glass envelope. The gas molecules inevitably come into contact with the cathode — positive ions are in fact attracted by it — and some of these molecules and ions combine with the free barium.

Cathode poisoning may also be due to another cause. Various adsorbed compounds — including evaporation products from the cathode — lie at the surface of electrodes, glass and mica. Electron bombardment releases some of these materials: of those reaching the cathode some combine with the free barium.

Another process occurring in the cathode, and which may lead to a valve failure, concerns the nature of the reducing agents in the cathode nickel. In the course of time the reducing agents react with the cathode coating at the boundary between the nickel tube and the coating. One reaction product may be a gradually thickening interface layer of barium orthosilicate (Ba<sub>2</sub>SiO<sub>4</sub>), which has a high resistance. The hotter the cathode, and the lower the current, the more rapidly the layer grows. This interface resistance effectively adds a resistor in the cathode lead shunted by a capacitance of the order of 10 000 pF; this not only reduces the cathode current but introduces negative feedback, causing a drop in gain, particularly at the lower frequencies<sup>7</sup>).

<sup>6</sup>) See for example, R. Loosjes and H. J. Vink, Philips tech. Rev. 11, 271-272, 1949/50.

<sup>7</sup>) From the extensive literature on this subject we may mention: A. S. Eisenstein, The leaky-condenser oxide cathode interface, J. appl. Phys. 22, 138-148, 1951, and M. R. Child, The growth and properties of cathode interface layers in receiving valves, Post Office Electr. Engr's J. 44, 176-178, 1951/52.

The behaviour of these different processes as a function of cathode temperature is very important. At temperatures of 50° to 100 °C below normal, the cathode is much more susceptible to poisoning. The reason is that the supply of free barium to the cathode surface is slowed down at low temperatures, as are diffusion processes in general, and also that certain reactions then take place between barium and other elements that do not occur at normal working temperatures. At higher temperatures, however, evaporation is more pronounced and the interface layer grows faster. There is therefore an optimum cathode temperature, which can be lower the more successfully cathode poisoning has been combated.

To obtain the best results certain precautions should be taken by the valve manufacturer as well as by the equipment designer (the valve user). We shall first consider the manufacturing side.

- a) To reduce *poisoning effects* the configuration of electrodes and screens should be such that no electrons can impinge on glass or mica. The components, moreover, should be kept scrupulously clean, requiring special procedures for cleaning and assembly, such as degreasing, boiling in distilled water, degassing by annealing in vacuum before assembly. Dust-free assembly, long pumping schedules and the use of high-quality getters are also necessary.
- b) The *cathode temperature rating* should be as low as possible without increasing too much the susceptibility to poisoning. (There must, of course, be some safety margin to allow for heater voltage fluctuations.)
- c) To combat *interface effects*, "passive" nickel should be used for the cathode tube, i.e. nickel containing approximately 0.03% Mg and not more than 0.01% Si (ordinary nickel contains up to about 0.1% Si). It takes longer with passive nickel to activate the cathode, but after 10 000 hours operation at normal cathode temperature there is still no measurable interface resistance.

The equipment designer, or user, can do a great deal to increase the useful life of a valve by observing the following precautions:

- a) *Heater voltage*. As a general rule the heater voltage should never be allowed to deviate from the nominal value by more than  $\pm 5\%$ . (For this reason, the heaters should not, where it can be avoided, be fed in series.) Where the heater voltage has been stabilized to within 1%, however, the valve can quite permissibly be run at, say, 5% (in some cases even at 10%) below the nominal value.

- b) *Bulb temperature*. To minimize the release of gas from the glass the bulb temperature should not be allowed to exceed 170 °C at the hottest point. In this connection, the use of screening cans round the valves calls for some care.
- c) *Anode and screen-grid dissipation*. High dissipation causes over-heating of the anode or screen-grid, or both, and also raises the temperature of the cathode, entailing the risk of gas formation and barium evaporation. With regard to poisoning it makes a difference whether a given dissipation is produced by a high voltage and a low current or by a low voltage and a high current. Poisoning effects are more pronounced at high than at low voltages, probably because the compounds adsorbed at the surfaces of anode and screen grid are more readily broken down by high-energy electrons into substances harmful to the cathode. It follows, therefore, that low dissipation at low voltage is advantageous; the permissible cathode current may quite be high.

#### *Insulation faults, grid emission and ion current*

Material evaporating from the cathode gives rise to faulty insulation by forming a conducting film on the mica spacers and the glass base of the valve. Deposits formed on the control grid can lead to grid emission. This cause of gradual failure may be remedied by using passive cathode material and by keeping the cathode temperature as low as possible and the control grid cool. The insulation over the surface of the mica spacers can be considerably improved by a coating of magnesium oxide.

When the grid is negative, positive ions in the valve cause some grid current to flow. In a well-manufactured valve with an efficient getter, however, so few positive ions reach the first grid per second that the measured grid current is only  $10^{-8}$  to  $10^{-9}$  A, of which a considerable fraction in any case arises from another cause, namely that the electrodes exposed to electron bombardment emit small quantities of soft X-rays, which release photo-electrons from the control grid. Genuine "ionic currents" (of the order of 1  $\mu$ A) are found only in valves with an air leak or with an inefficient getter. Both these defects are abnormal and can only be due to undetected manufacturing errors.

The manufacturer can prevent the occurrence of such defects in delivered valves by first keeping them in storage for a month and then checking for the presence of gas current. A very efficient method of checking for air leaks is to apply the *argon test*. If a valve is leaky, a good getter will absorb all

constituents of the intruding air except the argon (about 1%). The argon test consists in storing the valves for, say, twenty-four hours in a container filled with argon at atmospheric pressure; after this time the presence of a leak can easily be ascertained by measuring the grid current. In this way, storage for one day is equivalent to a shelf life of 100 days.

The equipment designer can often reduce the effects of imperfect insulation and grid emission by keeping the control-grid resistance as low as possible.

### Sudden failures

Sudden failures may be due to broken connections (e.g. a bad weld), short-circuiting between two electrodes, glass cracks and heater-cathode short-circuits.

### Broken connections

Practically all open connections in valves occur at or near welds. A reliable welding technique (spot welding) is therefore essential in the manufacture of valves. Non-destructive testing of every individual weld is obviously impracticable, but a generally high standard of welds can be obtained if the following rules are observed:

- 1) The valve designer should ensure that all welded connections are clearly visible and easily accessible, and that only readily weldable materials are used.
- 2) The welding equipment should yield a completely reproducible result.
- 3) Sampling tests on the quality of welds should be carried out at regular intervals.

The second point requires some elucidation. The quality of a spot weld depends on many factors, the most important being the pressure on the weld, the no-load voltage on the transformer secondary, the impedance of the whole secondary circuit, the phase at the moment the current is switched on, the time of the current cycle, the nature of the materials to be welded, the shape and surface of the electrodes, the mass of the moving parts and the atmosphere in which the joint is made. In a modern spot welder (*fig. 4*) all variables are accurately controlled. The optimum pressure for the usual welds is pre-adjusted to a fixed value found experimentally. The secondary voltages for each individual weld are specified in a table; in general, two or three voltages suffice. The operator places the parts to be welded in the appropriate position between the electrodes and starts the weld by depressing a pedal, after which the process is entirely automatic and free of the human element. The transformer primary current flows through two

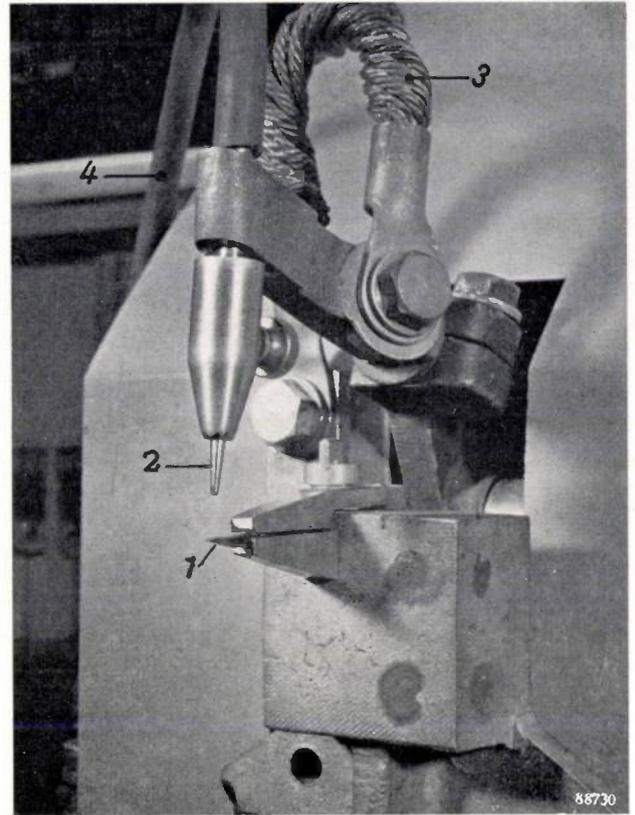


Fig. 4. Spot-welder. 1 fixed electrode, 2 moving electrode, 3 current supply cable, 4 hose through which a stream of a reducing mixture of  $H_2$  and  $N_2$  is played on the weld.

thyratrons in anti-parallel, which open and close the power circuit at the required times. The phase of the initiating current is fixed and entirely independent of the operator. The weld is terminated automatically, usually after one complete cycle of the mains voltage. (One cycle is normally sufficient, but in some cases several cycles may be required, which are here obtained by means of a control circuit using cold cathode valves.) The welding electrodes, made of a copper alloy, are exchanged at regular intervals, the old ones being reshaped on a special milling machine. A stream of a mixture of hydrogen and nitrogen is played on the weld to prevent oxidation. The mass of the moving arm is small, so that its inertia causes no significant variation of the electrode pressure as the materials compress during the welding operation. All these measures ensure reproducible welds of a consistently high quality.

### Short-circuits

Short circuits due to direct contact between two electrodes (which may be only 100  $\mu$  or even 50  $\mu$  apart) are easy to detect during production and are therefore extremely rare in the finished product. A much more difficult problem is the prevention of

partial short-circuits or conducting paths arising from foreign particles in the valve. "Short-circuits" of this nature may show values of resistance varying from  $10^7$  to 1 ohm. Moreover, they may only be intermittent, which makes them very difficult to detect.

These particles originate in many ways. In the first place they may be present in the air as dust, and therefore the air supply to the assembly shop is always filtered. The formation of dust in the assembly shop itself is reduced by using materials that give off little dust for walls, ceiling, benches, curtains, etc. Floors and benches should be repeatedly wiped down with wet cloths. Well-polished linoleum is a good dust-trap. The parts are assembled and welded under dust covers, screened off at the front by nylon curtains as shown in *fig. 5a* and *b*.

The components themselves, of course, must also be free of dust. They are therefore produced under conditions that are as dust-free as possible, after which they are subjected to washing and stoving processes before being conveyed, suitably packed, to the assembly shop. Only glass, metal or plastics can be used as packing material. For the conveyance and storage of grids, for example, a metal packing has been developed in which the grids can remain during the washing and stoving processes.

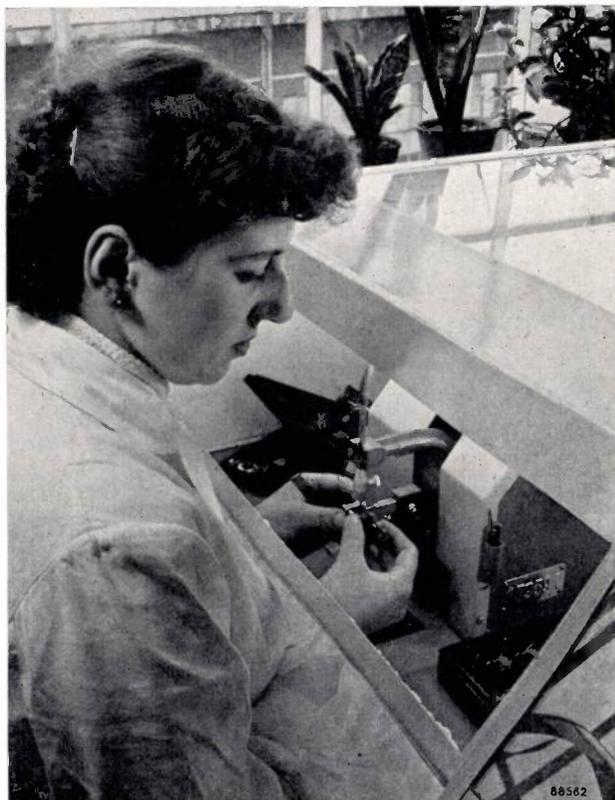


Fig. 5b. Spot-welding the electrode system of an amplifier valve.



Fig. 5a. Assembling the electrode system of an amplifier valve. This operation, like that in *fig. 5b*, is carried out under a transparent dust-cover. The operator's hands enter the dust-cover through slits in a nylon curtain, which produces no dust itself.

Finally there are the particles introduced during production: metal particles may fly off during welding, the coating of the cathode may be chipped and bits may flake off the mica spacers (mica particles become conductive in the long run under a coating of cathode material). Sputtered metal particles point to faulty adjustment of the welding machine, and chipped cathode material indicates a lack of care in the manufacture of the cathode. Formation of mica particles can be reduced by using a new form of spacer, which flakes much less than the old type (*fig. 6a* and *b*).

In spite of all these measures, the quality of a valve is still to some extent dependent upon the girl at the assembly bench. It is not enough to provide her with the best tools and to simplify and reduce the fatigue of the operations she has to perform. It is equally necessary to ensure that her working environment is fresh, clean and orderly as befits the quality expected of the product. What is more, the operatives should understand the importance of their work; this should be explained to them during their training period. Last but not least, the results are considerably influenced by the wage policy. In our opinion the best method is to pay a piece-rate in combination with a system of bonuses based on the quality of the work performed.

The quality is expressed in so many marks out of ten, the marks being given, in accordance with a fixed schedule, after making a number of sample tests.

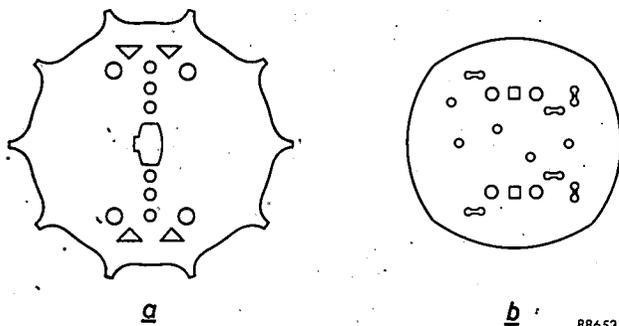


Fig. 6. Old form (a) and new form (b) of mica spacers. The new form reduces the incidence of flaking and chipping.

### Glass defects

After the electrodes have been mounted on metal rods, which are fused into a glass base, the glass base is sealed into the envelope. This is done in an automatic sealing machine, which must be checked constantly on its performance in order to prevent the occurrence of glass strains which might ultimately lead to cracks and hence to sudden failures. The variables here are the gas jets, the intensity and height of which depend upon the composition and pressure of the gas, and the glass, which is not a perfectly constant product. Each production run is preceded by the sealing in of a number of dummies, which are inspected for glass strain with the aid of a polariscope<sup>8)</sup>, and for the form of the seal by making a cross-section. After the machine has been adjusted for optimum sealing, the production run is started. Every hour samples are taken: after cooling, the sample valves are immersed in boiling water and a metal cone of prescribed form is pressed between the pins; they are then plunged directly into cold water. The occurrence of cracks gives indications of how the machine should be readjusted. After exhausting, all valves are immersed for a moment in boiling water. This test should result in no rejects; a reject is a sign that the sealing machine needs readjustment.

### Heater-cathode short-circuits

One type of failure that is due to a gradual physico-chemical process and yet occurs suddenly, without any warning, is the breakdown of the insulation between heater and cathode<sup>9)</sup>. The nickel tube of an indirectly heated cathode is heated to

a temperature of 750-800 °C by a tungsten (usually helical) filament, which itself has a temperature of about 1100 °C (fig. 7a). The only material that has been found suitable at this temperature to insulate the heater from the cathode, while preserving adequate adhesion and flexibility, is alumina ( $\text{Al}_2\text{O}_3$ ). The tungsten is coated with alumina by spraying or by electrophoresis<sup>10)</sup>, and afterwards sintered at a high temperature.

In most cases the heater-cathode insulation does not have to satisfy very stringent requirements. For example, in a high-frequency amplifier, with the heaters connected in parallel, the potential difference between the cathodes and the middle of the heater is only a few volts, viz. the voltage drop over the cathode resistor. It is quite another matter, however, in circuits where the heaters are connected in series or where the cathode is used as an output electrode and is at a high potential with respect to earth. There is then a danger that the insulation will ultimately break down owing to gradual electrolysis of the alumina at those points where it makes direct contact with the cathode nickel. When the heater is positive with respect to the cathode (the other polarity will be considered presently) negative oxygen ions are drawn to the heater where they cause oxidation. The tungsten oxide thus formed dissolves in the alumina, producing an aluminium tungstate (fig. 7b). The resistance of this compound is much lower than that of alumina. After a time determined by the voltage  $V_{kf}$  and the temperature of the heater the insulation between

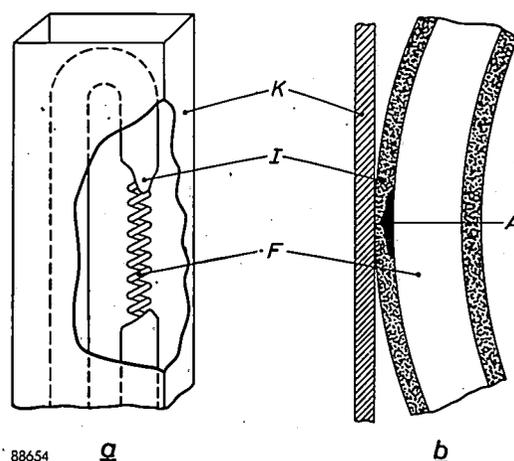


Fig. 7. a) Indirectly heated oxide cathode. K nickel cathode tube, F tungsten heater, surrounded by an insulating layer of alumina I.

b) If the heater is positive with respect to the cathode, a conducting layer of an aluminium tungstate (A) forms in the alumina coating at the points where the latter touches the cathode nickel, which may lead to a heater-cathode short-circuit. (Drawings not to scale.)

<sup>8)</sup> Philips tech. Rev. 9, 277-284, 1947/48; 14, 290, 1952/53.

<sup>9)</sup> This phenomenon has been investigated by P. G. van Zanten and P. N. Kuiper of the Physico-chemical Laboratory, Electronic valve factory, Eindhoven.

<sup>10)</sup> S. A. Troelstra, Philips tech. Rev. 12, 293-303, 1950/51.

the cathode and a point of the heater will suddenly break down, possibly resulting in a complete short-circuit. If the impedance in the heater-cathode circuit is low, the short may lead to the fusing of the heater.

Fig. 8 shows the relationship between  $V_{kf}$  (heater positive) and the time  $t$  in which 1% of a batch of valves develop heater-cathode shorts, with the heater voltage as parameter. (The values are for guidance only, having been obtained by extrapolation of accelerated tests carried out at

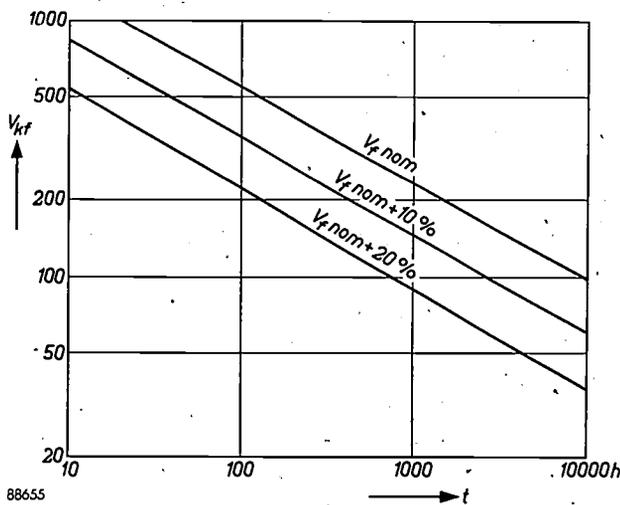


Fig. 8. Potential  $V_{kf}$  (heater positive with respect to cathode) as a function of the time  $t$  in which 1% of the valves show heater-cathode short-circuits, at nominal heater voltage  $V_{f\text{ nom}}$  (heater temperature 1080 °C) and at 10% and 20% higher heater voltages. Provisional curves only, obtained by extrapolation of accelerated tests.)

higher voltages and temperatures.) The figure shows clearly how the life of a valve is shortened by raising  $V_{kf}$  and how this is influenced by the heater voltage. It can be seen that  $V_{kf}$  should not exceed 100 V if a life of 10 000 hours is required.

The situation is less serious if the cathode is positive with respect to the heater. In this case it is the nickel that is oxidized, but since nickel is not so readily oxidized as tungsten, the process takes much longer (about ten times longer). This means, however, that to obtain the same life as with the other polarity the value of  $V_{kf}$  may only be about twice that of the former case.

To counteract this phenomenon the manufacturer can try to prevent contact taking place between the cathode and the heater insulation touching the inside of the cathode, or at least to allow contact only via a long insulation path. This is done for instance in booster diodes for television receivers, in which peak voltages of many thousands of volts occur between cathode and heater. However, the method is only practicable for large cathodes, and

has the drawback of requiring a higher heater temperature, which tends to make heaters brittle. In many Philips valves for professional use the method adopted for smaller cathodes is to coat not only the tungsten but also the inside of the cathode tube with a layer of alumina. This has two advantages:

- The temperature of the inside coating is lower than the coating on the tungsten, which offers a greater safeguard against breakdowns.
- The heat transfer is improved by the higher radiation absorption factor of alumina compared with nickel, which means that the heater temperature can be lowered.

It has been found that valves with cathodes treated in this way (types E80CC and E80L are examples) function 5 to 10 times longer before insulation breakdowns occur between heater and cathode, or that, for the same life,  $V_{kf}$  can be appreciably higher.

Insulation breakdown due to electrolysis currents as described above is in no way connected with the leakage currents, often much larger, that may flow between heater and cathode. Such currents may be due to emission phenomena (e.g. the presence of BaO on the tungsten or on the inside of the cathode) or to ionic conduction arising from impurities (e.g. sodium) in the alumina coating.

#### The contacts between valve and valve holder

The quality of base pins and socket contacts, by which a valve is connected with the circuit, has a very important bearing on the quality and reliability of the whole. An investigation conducted by Morrell<sup>11)</sup> of the British General Post Office demonstrated that changes in level (even interruptions) and noise in long-distance telephone communications are due in large measure to bad contacts, including base-pin contacts. Morrell concludes that the only answer is to solder the valves into the circuit.

In our opinion this drastic step is acceptable only where the life of the other circuit elements is of the same order as the valve life, or shorter. Usually, however, equipment life is many times longer than the average life of the valves, so that the replacement of soldered-in valves would create a serious maintenance problem. On the whole, new valves and new valve-holders make excellent contact; the contact deteriorates in the course of time owing to corrosion. Only precious metals maintain stable contact under corrosive conditions, a fact which has been proved by the excellent results obtained from a series of extensive tests car-

<sup>11)</sup> F. O. Morell, paper read before the London branch of the Institution of Post Office Electrical Engineers, 5 April 1948.

ried out, under very adverse conditions, with gold-plated pins and socket contacts. The majority of Philips "Special Quality" valves are now equipped with gold-plated base pins.

### Spread in characteristics

Discrepancies in electrical characteristics between valves of the same type are attributable to differences in the dimensions or clearances of electrodes, in the work function of the control grid and in cathode emissivity.

Pentode type 18042 will serve to illustrate the effect of dimensional variations on the electrical characteristics. This valve has clearances of 120  $\mu$  between cathode and first grid and 310  $\mu$  between the first and second grids. The following table shows the change in grid voltage  $V_{g1}$  and mutual conductance  $S$  at 10 mA anode current (screen-grid potential 120 V) for 1 $\mu$  variations of various electrode dimensions.

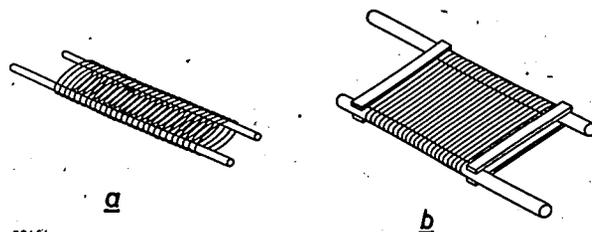
	Change in $V_{g1}$ mV/ $\mu$	Change in $S$ (mA/V)/ $\mu$
Cathode diameter	10	0.06
Grid 1 diameter	1.5	0.06
Grid 1 pitch	90	—
Grid 1 wire diam.	115	—
Grid 2 diameter	5	—

It is evident that special measures are needed in the manufacture of this type of valve in order to satisfy the specification, say, that  $V_{g1}$  should not vary by more than  $\pm 330$  mV from the nominal value (as required for the pentode 18042). It calls, among other things, for extremely narrow tolerances in the diameter and tensile strength of the wire for the first grid, and for the utmost accuracy in the cathode diameter (the grid pitch can be set precisely relatively easily).

The work function of the first grid in a new valve may fluctuate initially, but will settle down if the valve is allowed to burn in for 48 hours. The work function depends largely upon the extent to which evaporation products from the cathode have become deposited upon the grid surface, that is to say it depends on the cathode temperature. The grid is more susceptible to poisoning than the cathode since it does not itself produce free barium but can only receive it from the cathode. The aim of the 48 hour ageing period is therefore to allow the surface of the grid to reach its normal working condition, after which there will be no appreciable change in the work function.

In discussing electrical tolerances, mention should

be made of a new grid construction, evolved elsewhere<sup>12</sup>). Until quite recently, grids were almost invariably made as shown in *fig. 9a*. Molybdenum grid wire is wound helically around two "backbones" made of nickel, copper-clad wire or a similar material, in which grooves are cut to hold the wire turns in place, the grooves being closed under pressure after winding is completed. The grid wire itself, whose turns may be oval or round in form, have not only an electrical function but also a mechanical function, in that, together with two mica spacers, they determine the spacing between the backbones. In the new construction<sup>12</sup>), illustrated in *fig. 9b*, the grid



88656  
Fig. 9. A grid of normal construction (a) and a modern "frame grid" (b).

wires have a purely electrical function. The mechanical function is fulfilled by a sturdy frame consisting of two molybdenum rods connected, with the right separation by four spot-welded molybdenum strips. Very thin tungsten wire (e.g. 7.5  $\mu$ ) is wound tightly around the frame, the winding tension corresponding to about 60% of the tensile strength of the wire. The wire is fixed to the frame with gold solder or with molten glass powder. The thickness of these frame grids, that is to say the clearance between the turns on opposite sides of the frame, is wholly determined by the thickness of the molybdenum rods, which can be manufactured within a precision of 5  $\mu$ .

In conjunction with measures to reduce the tolerance in the dimensions of the cathode and the mica spacers, the use of a frame grid allows valves to be made with a cathode-grid clearance of 50  $\mu$ , that is, half the clearance that can be reached satisfactorily with normal grids. Several types of Philips valves have already been fitted with frame grids (e.g. E180F and E88CC, shown in *fig. 1*) and other types will follow in the near future.

### Shock and vibration resistance

The enormous increase in the use of electronic equipment in aircraft has made it essential to make valves that can withstand severe vibrations and shock.

<sup>12</sup>) G. T. Ford, Bell Labs, Record 27, 59-61, 1949. E. J. Walsh, *ibid.* 28, 165-167, 1950. G. T. Ford and E. J. Walsh, Bell Syst. tech. J. 30, 1103-1128, 1951.

If there is the slightest play between the mica spacer and the glass envelope, or between the mica and a component which it holds in place, the mica will begin to wear as soon as the valve is exposed to vibrations and this will tend to increase the amount of play. The first of several harmful consequences will be increased microphonic noise. In an advanced stage the play may cause shorts between two or more electrodes. Moreover, the mica may release gas, and mica flakes may come in contact with the cathode and poison its surface. Mechanical fatigue may lead to rupture, particularly of welds, filament and cathode connecting strip.

To improve a valve's ability to withstand adverse mechanical conditions it is first of all necessary to restrict the forces acting on the components. This is done by using the lightest possible components and by avoiding mechanical resonances under 200 c/s, and preferably under 1000 c/s. The toothed mica spacers formerly in general use (fig. 6a) are liable to break off at the teeth under vibration (especially in large valves) which gives rise to play and to flaking mica. The new form of spacer illustrated in fig. 6b is a substantial improvement. An especially good fit is obtained by very accurately narrowing the bulb at the height of the spacer to an inner diameter about 0.1 to 0.2 mm smaller than the largest diameter of the spacer. Fig. 10 shows the interior of a valve in which this method has been employed. Mechanical fatigue of the heater can be prevented by restricting the movement of this element inside the cathode tube. The cathode connector can be safeguarded against rupture by giving it a large bend.

Various methods have been devised for testing valves under adverse mechanical conditions. The American military specifications, which are used as standard for Philips valves, prescribe a vibration test lasting in total 96 hours, divided into three equal periods of 32 hours, during each of which the acceleration force is applied in one of the three main axes of the valve. The specified repetition frequency is 25 c/s and the peak acceleration 2.5 g (g being the acceleration due to gravity), and the amplitude 1 mm. The American specifications also lay down a shock test consisting of five shocks of 500 g, each lasting 1 millisecond, and applied in the direction of each of the axes; for the longitudinal axis five blows are given in the one direction and five in the opposite direction. This test is carried out on a machine of American design. (fig. 11).

The valves are considered to have passed these tests when they are still in working order and when their electrical characteristics still fulfil certain requirements. As the tests are, in a sense, of a destructive nature, they can only be carried out by sampling. The American military specifications therefore lay down sampling plans<sup>13)</sup> which are based on an "acceptable quality level" of 6.5% rejects for the vibration test and 20% for the shock test. The figures indicate that valves which have satisfied these specifications should not be exposed in practice to vibrations and shocks of the same magnitude as during the tests, as otherwise the failure rate would be very high.

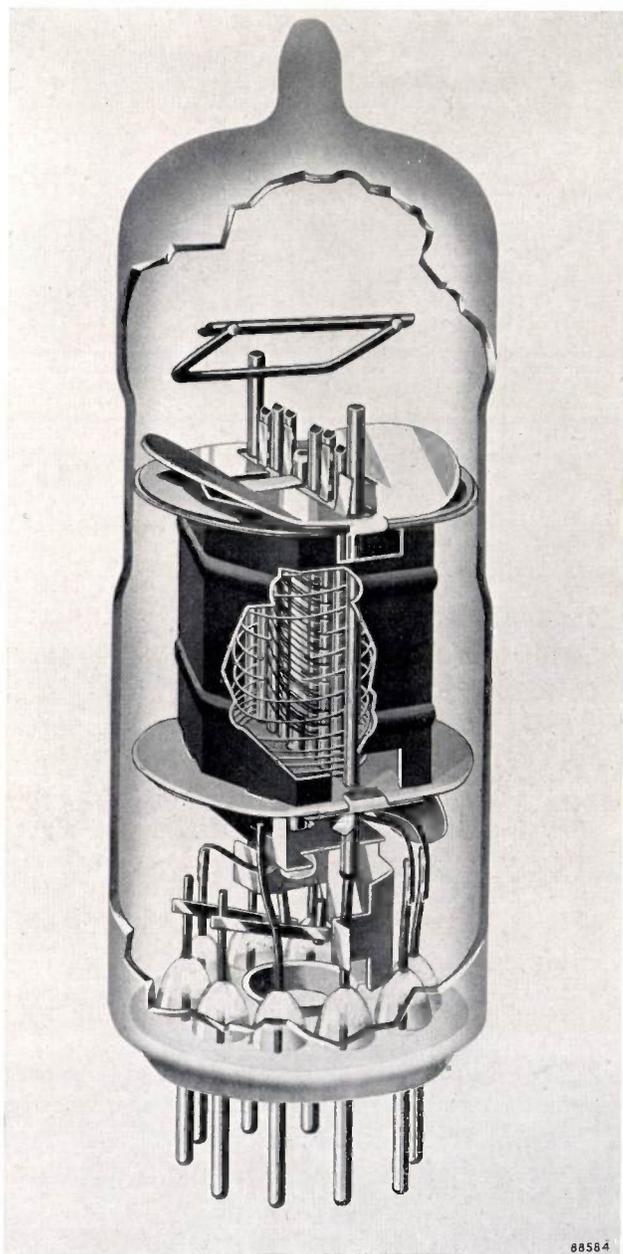


Fig. 10. Type E80F pentode, fitted with mica spacers as in fig. 6b. The upper spacer fits tightly in a constricted region of the envelope.

<sup>13)</sup> For a general treatment of sample testing and sampling inspection plans, see Philips tech. Rev. 11, 176-182, 260-270 and 362-370, 1949/50.

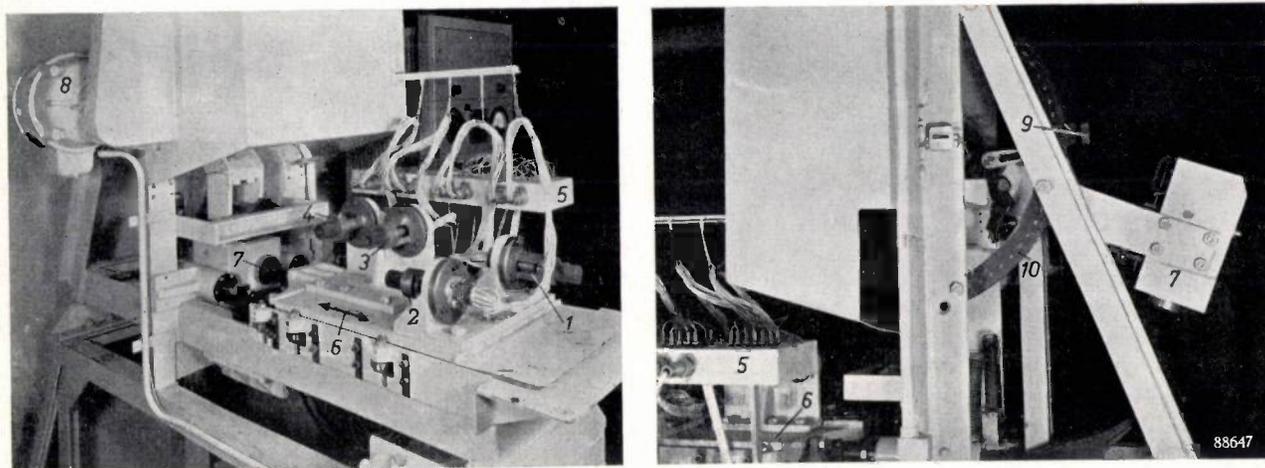


Fig. 11. Two views of a machine for shock-testing valves in accordance with American military specifications. 1, 2, 3, and 4 are valve bases mounted facing different directions, connected to the supply rack 5, and fixed rigidly to a carriage 6 which can slide in the directions indicated by arrows. A hammer 7, driven by a motor 8, strikes the carriage after a fall from a height adjusted by screw 9 along a scale 10. After 5 consecutive blows the positions of the valves are changed until they have passed through all four positions 1 . . . 4. They are then removed and inspected for damage.

### Quality control

In the manufacture of "Special Quality" valves, the methods of quality control play a very important part. Some examples have already been mentioned in connection with assembly work and glass strains. However, these methods start as far back as in the manufacture of components, where samples are tested on their most important properties. After pumping and "screening" (cathode activation and stabilization), the valves are tested for short-circuits, open connections, crackling, etc., and certain of their electrical characteristics are measured. They are then subjected for five minutes to a vibration test in order to detect the presence of loose particles, and this is followed by a 48 hours' ageing period to stabilize the electrical characteristics, which constitutes a short life test on every valve produced. Rejects occurring during this period give valuable indications concerning the reliability of the valves. After ageing, all valves are tested on their more important characteristics, such as emission, mutual conductance and anode current and examined again for crackling and insulation. Each week's production is stored separately. From each batch several samples are taken, some being subjected to measurements on capacitances, grid emission, noise, microphony, etc. and others to a life test of 500 hours.

The results of these sample tests determine whether the week's production is suitable for delivery. Before dispatch, the valves are tested once again on their principal characteristics in order to ascertain whether they have suffered during storage.

### Conclusion

The purpose of this article has been to give some idea of the problems confronting the manufacturer of special quality valves for professional equipment, and also to indicate how the valves can be employed to best advantage in electronic equipment. As stated, a failure rate of 0.5% per thousand hours with a useful life of 10000 hours does not mean that endeavours to improve the reliability of these valves are at an end. Much work undoubtedly remains to be done before the failure rate can be reduced to 0.1% per 1000 hours. The equipment designer can help in reaching this goal by making available the extensive practical data at his disposal; close cooperation in this respect between valve manufacturer and valve user can do a great deal to improve the quality of this class of valves.

**Summary.** The greatly increased use of valves for professional equipment has focussed the attention of manufacturers and users alike on the reliability and life of these components. The authors define "reliability" as the reciprocal of the failure rate, i.e. the percentage failures in a batch of valves after 1000 hours operation. They regard the end of a valve's useful life as the moment at which the failure rate begins to increase. A useful life of 10000 hours and a failure rate of 0.5% per 1000 hours have already been reached with "Special Quality" type valves; it is hoped to be able to reduce the failure rate to 0.1% per 1000 hours.

Valve failures may be either gradual or sudden. The causes of both categories of failure are discussed separately and a description is given of the measures taken in recent years to reduce the incidence of valve defects. The spread of characteristics and shock and vibration resistance are also discussed. In conclusion the desirability of close cooperation between valve manufacturer and valve user is underlined.

# A FLYING-SPOT SCANNER FOR TELEVISION 35 mm FILM

by F. H. J. van der POEL.

621.397.611.2:778.53

*Flying-spot scanners are particularly suited for the televising of flat, transparent pictures. In a previous article in this Review a flying-spot scanner for televising stationary objects was described. A similar scanner has now been developed for the transmission of 35 mm cinefilm, and the present article discusses the problems involved in scanning a succession of images.*

## Introduction

The transmission of films has become a feature of every television programme. The equipment used for this purpose is required to derive an electrical signal, suitable for television, from standard film made for cinematographic projection. Such films are recorded at 24 frames per sec; the European television standards, however, require 25 pictures per sec. Since there is no simple method of obtaining 25 television pictures per sec. from 24 film frames per sec, the obvious expedient is to match the film speed to the speed required for television. Increasing the film speed from 24 to 25 frames per sec has no objectionable effect either on the picture or on the sound. In this respect Europe is better off than the United States and other countries where the mains frequency is 60 c/s, and where the number of television pictures per second is accordingly fixed at 30; the film speed cannot be raised so high without making movements appear too fast and music and speech too high-pitched. In their case, therefore, the solution is more complicated.

Every television picture is formed from a raster consisting of a certain number of horizontally scanned lines. According to the Gerber standard<sup>1)</sup> which is used for example, in the Netherlands, the number is 625. Moreover, the scanning lines are interlaced, which means that the odd lines 1, 3, 5, etc. are traced first, and then the even lines 2, 4, 6, ..... The vertical deflection of the light-spot which traces a television raster is shown in *fig. 1a* as a function of time. Each complete picture period  $T$  of 40 milliseconds ( $1/25$  sec) may be broken down into four intervals: the raster flyback times  $T_{b1}$  and  $T_{b2}$  are each 1.5 msec, leaving 18.5 msec each for the actual scanning periods  $T_{s1}$  and  $T_{s2}$ . Strictly speaking, this means that the number of lines per odd or even raster is not exactly  $625/2$ , but only  $(18.5/40) \times 625$ .

For the televising of films, frequent use is made of flying-spot scanners. The interlacing, particularly, gives rise to a number of problems. In the following description of a flying-spot scanner for 35 mm film developed in Philips Research Laboratories, the solution of these problems will be discussed. Before this we shall first touch briefly on a system of televising films with the aid of a normal television camera.

## Televising of films using a television camera tube

A familiar method of televising films is that in which a normal storage-type camera tube (iconoscope, image iconoscope) is used in conjunction with a slightly modified conventional cinema projector. The film is fed intermittently through the projector in the normal way, but at a rate of 25 frames per second. Each film frame is projected on to the light-sensitive layer in the camera tube during the brief flyback periods  $T_{b1}$  and  $T_{b2}$  (*fig. 1b*). The potential

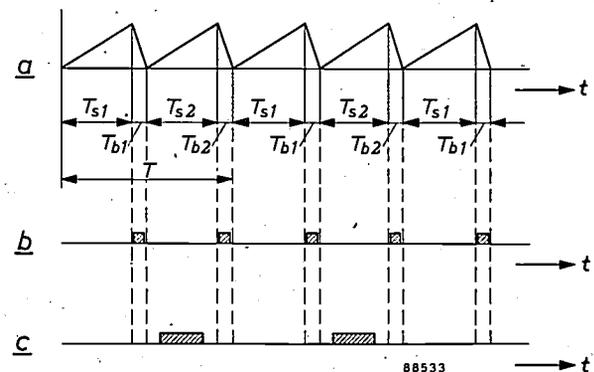


Fig. 1. a) Vertical deflection, as a function of time  $t$ , of the spot tracing a television raster.  $T$  = picture period (40 msec). The "odd" and "even" lines of the raster are scanned during the intervals  $T_{s1}$  and  $T_{s2}$  respectively (each 18.5 msec). During the flyback intervals  $T_{b1}$  and  $T_{b2}$  (each 1.5 msec), when the spot returns from the end point of one raster to the starting point of the next, the spot is suppressed.

b) Times during which the film frames are projected when televising a film with a normal television camera tube. These times fall within the intervals  $T_{b1}$  and  $T_{b2}$ .

c) The film is moved forward one frame during the  $T_{s2}$  intervals.

<sup>1)</sup> See, e.g., W. Werner, The different television standards considered from the point of view of receiver design, Philips tech. Rev. 16, 195-200, 1954/55.

pattern thus produced on the target of the camera tube is then scanned during the intervals  $T_{s1}$  and  $T_{s2}$  by an electron beam<sup>2)</sup>. During scanning the projected light beam is intercepted by a shutter. In this way use is made of the camera tube's "memory", in which the potential pattern remains for a while after the exposure has ended.

The advantage of this system, in that it allows a normal studio television camera to be combined with a conventional cinema projector, is offset by the fact that storage-type camera tubes give rise to signals even when not illuminated<sup>3)</sup>. During illumination these spurious signals are superimposed on the desired signal. Spurious signals are not troublesome if there is sufficient intensity of illumination, but this is by no means always the case in the televising of films.

With the development of camera tubes of the vidicon type, in which the drawback of spurious signals has been largely overcome, the system employing camera tubes has recently been gaining ground<sup>4)</sup>, particularly since the equipment needed is cheaper. At the present time, however, in Europe at least, flying-spot scanners are in general use and generate a television signal of very good quality.

#### Principle of the flying-spot scanner

As described in a previous article<sup>5)</sup>, a flying-spot scanner generates a well-defined and reproducible television signal from a flat, transparent object. The principle is briefly as follows.

A point source of light of constant intensity, formed on the screen of a specially designed cathode ray tube (the scanning tube) is projected by an optical system on to the transparent object — the film. (In the following we shall denote the point of light on the scanning tube as the *source* and its projection on the film as the *spot*.) The source is made to trace on the scanning tube a raster of the required number of lines, so that its image, the spot,

traces an identical raster on the film. The light passing through the latter is concentrated on a photoelectric cell by means of a condenser lens, so that a photo-current is generated which at every moment is proportional to the transparency of the film at the point defined by the spot. The current passes through a resistor, the voltage across which — after amplification — represents the required picture signal.

#### The film movement

A fundamental difference between the method of televising films with a flying-spot scanner and the method employing a television camera tube is that in the former it is the film image itself which is scanned, whereas in the latter it is the potential pattern formed in the camera tube. This means that, with a flying-spot scanner, the film image must be available for scanning during both intervals  $T_{s1}$  and  $T_{s2}$ ; it is not possible to use the interval  $T_{s2}$  for moving the film one frame farther, as is done in the camera tube method (fig. 1c). A process might be imagined in which the film is transported during the flyback interval  $T_{b2}$  and kept stationary for the rest of the picture period. In that case, however, a film frame would have to be moved forward in 1.5 millisecc, which would entail considerable mechanical difficulties. (In cinema projection it takes about 10 milliseconds to move a film one frame forward.) It has not yet proved possible to construct a 35 mm projector capable of such rapid film transport; even for 16 mm film, where the accelerations are so much smaller, the problem has not yet been satisfactorily solved.

However, it is not a prerequisite that the film should remain stationary during scanning. What is essential is that there should be a certain definite relative movement between the film and the spot, which can equally well be effected partly by the film and partly by the spot. In the flying-spot scanner to be discussed the film is transported at a *constant* speed, corresponding to 25 frames per second. We shall assume that the film velocity  $v$  is directed vertically downwards. The spot is made to move such that during each scanning period it has a velocity component  $w$  directed vertically upwards. Film and spot then have a relative vertical velocity of  $v + w$ . In a scanning period  $T_s (= T_{s1} = T_{s2})$  the spot has therefore moved over a distance  $(v + w) T_s$  with respect to the film. The value of  $w$  is chosen such that

$$(v + w) T_s = h \quad \dots \quad (1)$$

in which  $h$  is the height of the part of the film frame

<sup>2)</sup> See P. Schagen, H. Bruining and J. C. Francken, The image iconoscope, a camera tube for television, Philips tech. Rev. 13, 119-133, 1951/52. and J. C. Francken and H. Bruining, New developments in the image iconoscope, Philips tech. Rev. 14, 327-335, 1952/53.

<sup>3)</sup> See page 123 of the article by Schagen, Bruining and Francken referred to under <sup>2)</sup>.

<sup>4)</sup> H. N. Kozanowski, Vidicon film-reproduction cameras, J.S.M.P.T.E. 62, 153-162, 1954; R. G. Neuhauser, Vidicon for film pick-up, J.S.M.P.T.E. 62, 142-152, 1954; W. L. Hurford and J. Marian, Monochrome vidicon film camera, R.C.A. Review 15, 372-388, 1954; also L. Heyne, P. Schagen and H. Bruining, An experimental photoconductive camera tube for television, Philips tech. Rev. 16, 23-25, 1954/55.

<sup>5)</sup> A flying-spot scanner developed in Eindhoven is described by F. J. H. van der Poel and J. J. P. Valetton, The flying-spot scanner, Philips tech. Rev. 15, 221-232, 1953/54 and A. Bril, J. de Gier and H. A. Klasens, A cathode ray tube for flying-spot scanning, Philips tech. Rev. 15, 233-237, 1953/54.

to be scanned. The film is threaded in such a way that at the beginning of a scanning period  $T_{s1}$  (fig. 1a) the spot starts at the bottom of a film frame, and then scans the odd lines during this period (fig. 2a and b). After the flyback period  $T_{b1}$  the spot begins to scan the even lines, which should fall exactly between the lines scanned during  $T_{s1}$ . However, as may be seen from fig. 2c, this is not the case; line 2, for example, does not fall between lines 1 and 3 but half a frame higher. This spot is therefore intercepted by a shutter during the scanning of the even raster. How the even

The distance  $s$  between two film frames (the frame pitch) being 19 mm, the film velocity  $v$  is equal to  $25 \times 19$  mm/sec. The standard height and width of the part of the film frame to be projected are 15.2 and 21 mm respectively, and their ratio is therefore 3 : 4.14. The ratio needed for a television picture is 3 : 4, so that to avoid showing on the television screen a part of the film frame not intended for projection, the whole height but not the whole width of the film image should be scanned. For  $h$  in (1) we may therefore insert 15.2 mm, and since  $T_s = 18.5 \times 10^{-3}$  sec, we can now calculate  $w$ . The height of the raster in the film plane then follows from (2) as  $h' = 6.4$  mm. The width must be  $\frac{4}{3} \times 15.2 = 20.3$ , which means that the ratio of height to width is not 3 : 4 but 1.26 : 4.

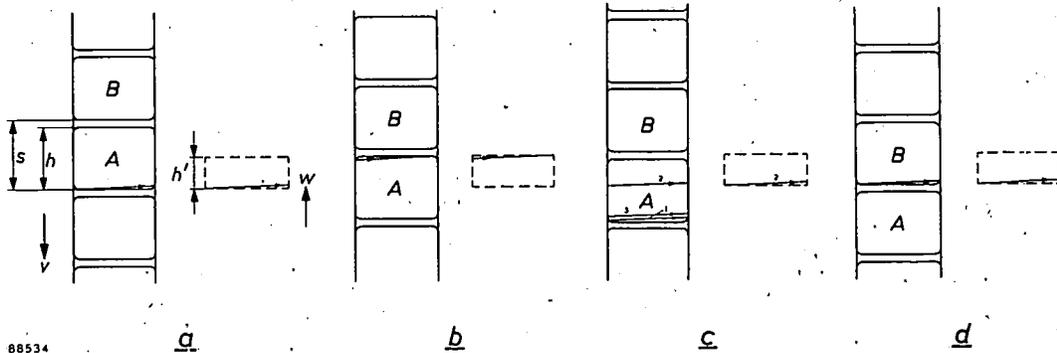


Fig. 2. a and b) The situations at the beginning and end of an "odd" raster scanned on film frame A. For clarity the outline of the raster is shown at the side of the film.  $v$  = film velocity;  $w$  = vertical velocity of scanning spot;  $s$  = distance between two frames;  $h$  = height of part of film frame to be scanned;  $h'$  = height of raster in film plane. c) For the beginning of the "even raster, the scanning spot starts at the wrong position on the film (line 2 should fall between 1 and 3); it therefore has to be cut off. d) After a full picture period the situation is again as in (a), but the film has moved forward one frame.

raster is actually produced will be described in the next section. The scanning of the odd lines, however, presents no further problems; when the spot begins,  $\frac{1}{25}$  sec later, to trace the odd lines again, the film has moved exactly one frame farther (fig. 2d).

Since the film movement provides for part of the relative vertical movement of the spot, the latter need not traverse the entire height  $h$ , so that the height  $h'$  of the raster in the film plane will also have to be smaller than  $h$ . (The film plane, i.e. the plane in which the film moves during the scan, should be clearly distinguished from the film itself: the film is in motion whereas the film plane is stationary.) The height  $h'$  can easily be calculated. Since  $v$ ,  $T_s$  and  $h$  are known, the value of  $w$  follows from (1) and  $h'$  from

$$h' = wT_s \dots \dots \dots (2)$$

Since  $h' < h$ , the ratio of height to width of the raster scanned by the spot will be smaller than the 3 : 4 prescribed for a normal television raster. The same obviously applies to the raster traced by the source on the scanning tube.

**Interlaced scanning**

Since the film moves at a constant speed, it will have travelled half the frame-pitch during the time  $T_{s1} + T_{b1}$  (20 msec) that elapses between the start of an odd raster and the start of an even one. In the film plane, therefore, the envelope of the even raster should not coincide with that of the odd raster but should be displaced over half the frame pitch ( $s/2 = 9.5$  mm) in the direction of film travel. If this is done the lines of the two rasters will be correctly interlaced on the film.

The requisite displacement of the even raster is effected in the present equipment by projecting a second image of the light source on to the film by means of a second objective mounted below the first, in such a way that the distance between the two images is equal to exactly half the frame pitch (fig. 3a). A rotating shutter, directly behind the film, intercepts the beam from the lower objective during the periods  $T_{s1}$  and that from the upper objective during the periods  $T_{s2}$ . The requirements as regards the height and speed of the shutter blades can be seen from fig. 4, which shows

as a function of time the vertical deflections of the upper and lower spots as well as the vertical positions of the upper and lower edges of a succession of shutter blades. The regions between these latter

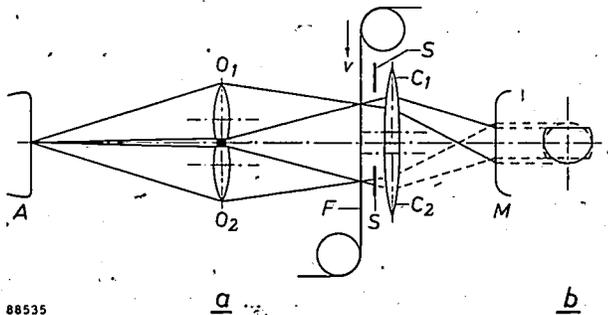


Fig. 3. a) To ensure that the even raster will interlace correctly with the odd, two images of the source on the scanning tube *A* are projected on to the film *F* by two identical optical systems *O*<sub>1</sub> and *O*<sub>2</sub>, mounted one above the other. Two condenser lenses *C*<sub>1</sub> and *C*<sub>2</sub> collect the light passing through the film and concentrate it upon the photo-electric cell *M*. A rotary shutter *S* (not to scale) between the film and the condenser lenses alternately intercepts one of the light beams (the lower one in the figure). *C*<sub>1</sub> and *C*<sub>2</sub> project the images of *O*<sub>1</sub> and *O*<sub>2</sub> respectively upon the light-sensitive layer of the photo-electric cell *M* and thereby spread the light from the spots over the whole images.

b) The images of *O*<sub>1</sub>' and *O*<sub>2</sub>' projected on the photo-electric cell by *C*<sub>1</sub> and *C*<sub>2</sub> respectively (the image of *O*<sub>2</sub>' is drawn with dashed lines). *C*<sub>1</sub> and *C*<sub>2</sub> are so spaced as to make the images coincide as closely as possible.

pairs of lines are hatched; the distance between them, measured perpendicular to the abscissa, represents, of course, the height of the blades. It is assumed that the blades are moving vertically upwards at a constant speed, an assumption which is, of course, only approximately true, for the shutter radius is not infinitely large. The slope of the lines that

represent the upper and lower edges of the shutter blades is a measure of the vertical velocity of the blades. Blade velocity and blade height should be so chosen that during all *T*<sub>s1</sub> periods the sawtooth for the lower spot falls within the hatched area, while the sawtooth for the upper spot falls outside it. During all *T*<sub>s2</sub> periods the reverse should be the case. It can at once be seen from the figure that it is possible to satisfy these conditions with many different combinations of blade height and blade velocity. Since it is just as important to transmit the whole of the active beam as it is to intercept the other, it is logical to make the distance between successive blades equal to the height of the blades. Fig. 4 represents the situation obtaining in the flying-spot scanner designed in this laboratory.

Complications arise from the finite radius of the shutter; for example, at the inner radius the blade velocity is less than at the outer radius. However, the margin allowed is sufficient to compensate for this discrepancy, as well as for the fact that the beams of light have a larger cross-section at the position where they are intercepted than at the position of the spots themselves (fig. 3a).

A relation exists between the number of shutter blades *n*, the radius of the shutter (for which we take the radius *r*<sub>m</sub> to the middle of the blade) and the velocity *v*<sub>m</sub> of this midpoint. In 1/25 sec the shutter must have moved exactly one blade farther, i.e. it must have turned about an angle of 2π/*n* radians. The angular velocity of the shutter must therefore be ω = 25 × 2π/*n* rad/sec. Since *v*<sub>m</sub> = ω*r*<sub>m</sub>,

$$v_m n / r_m = 50 \pi \dots \dots \dots (3)$$

It can be seen from fig. 4 that the widest margin will be obtained if the hatched band is parallel to the line *AB*. In that case the blade velocity is equal to the film velocity (for the vertical distance between *B* and *A* is 9.5 mm, i.e. half the frame pitch, and the horizontal distance is equal to half a picture period). This value of blade velocity is accordingly taken for *v*<sub>m</sub>. The value of *r*<sub>m</sub> is fixed by (3) after the number of blades *n* has been decided on; *n* cannot be made arbitrarily small as otherwise *r*<sub>m</sub> will also be very small and the velocities on the inner and outer edges of a blade will deviate too much from *v*<sub>m</sub>. In the present case, *n* is 24, which makes *r*<sub>m</sub> = 73 mm and the outer radius 85 mm.

The two light spots in the film plane are given the requisite spacing of half the frame pitch (*s*/2) by adjusting the vertical distance between the axes of the two objectives. Since there is always a certain amount of shrinkage in a film, depending upon age, humidity and temperature, the frame pitch is not precisely fixed, being invariably slightly less (about 1% max) than the nominal 19 mm. The distance between the objectives should therefore be adapted to the shrinkage in the film before beginning a transmission. Experience has shown

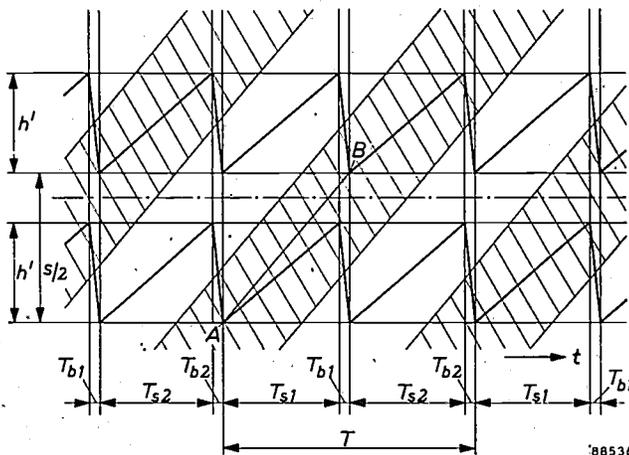


Fig. 4. Graphical representation of the alternate transmission and interception of the two scanning spots by the shutter blades. The two sawtooth lines represent the vertical deflections of the upper and lower spots as a function of time *t*. The boundaries of the hatched regions indicate the top and bottom edges of successive blades as functions of the time. In the intervals where a sawtooth falls within a hatched area the beam concerned is intercepted. Meaning of the symbols as in figs. 1 and 2.

that the amount of shrinkage in any one film does not change perceptibly over the whole length, but may vary considerably between different films. If the vertical distance between the objectives is not correctly adjusted, the lines of the two rasters on the film will no longer be properly interlaced. This does not, it is true, affect the interlace at the receiving end (where, of course, the television receiver is itself responsible for the interlacing on the picture tube) but the parts of the picture carried by the two rasters of the flying-spot scanner are no longer complementary. This is visible as a vertical

other along the horizontal. This *horizontal* doubling of the picture can be avoided if the optical systems are mounted with proper care.

The central section of the film scanner is shown in the photograph in *fig. 5*.

Other, perhaps more obvious, methods are conceivable for projecting the required images of the odd and even rasters in the film plane. It would be possible, for instance, to arrange for the two rasters to be traced at different positions on the scanning tube and then projected, by only one objective, in the correct relative positions on the film plane. It would then be necessary to ensure that the displaced even raster on the scanning tube is exactly complimentary to the odd raster.

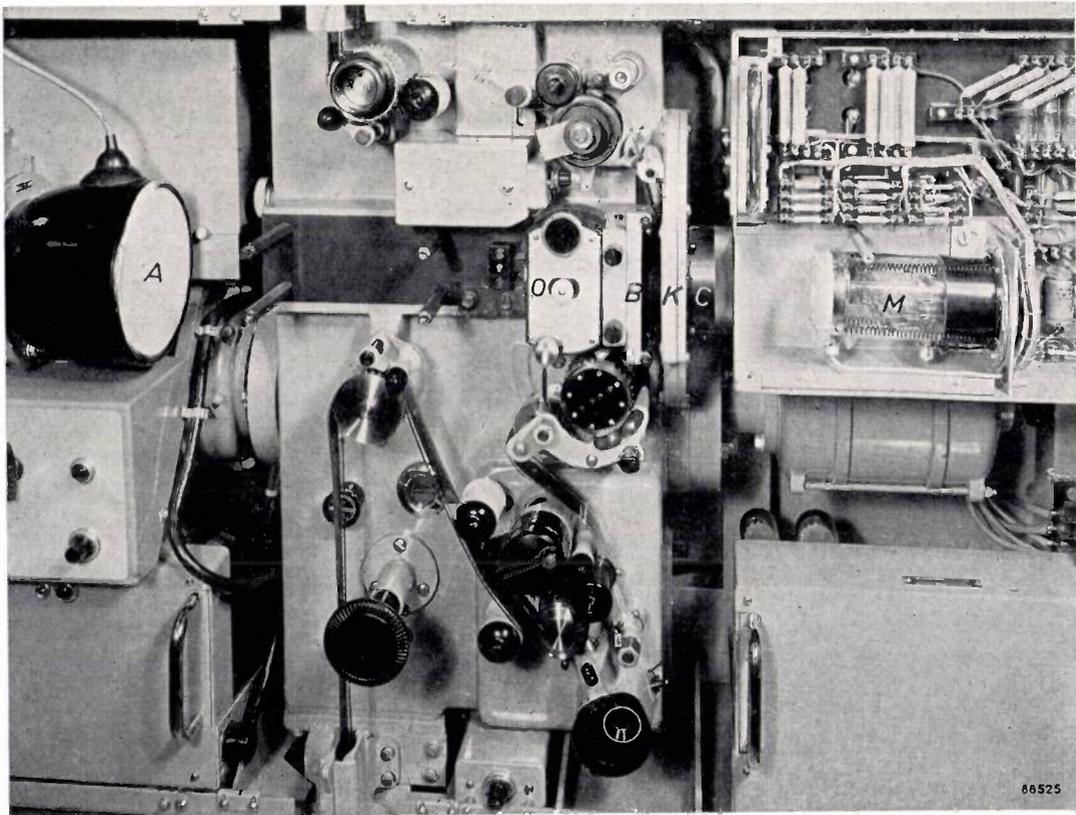


Fig. 5. The central section of the flying-spot scanner, showing the film in position. *A* scanning tube, *O* mount for double objective, *B* film gate, *K* shutter housing, *C* mount for condenser lenses, *M* photo-electric cell.

doubling of the picture, a phenomenon which serves as a criterion for judging whether the distance between the objectives has been correctly adjusted.

If the line joining the centre points of the objectives is not exactly parallel with the direction in which the film is moving, the two projected images of the source will be displaced with respect to each other at right-angles to that direction. The points of a vertical line on the film, produced by the odd and even rasters respectively, will therefore appear on the picture tube displaced with respect to each

This is in fact impracticable, however, for with two non-centrally disposed rasters, it presupposes the strictest accuracy in the deflection of the source on the scanning tube. The method used in the present equipment imposes the requirement of strict accuracy upon the objectives, which are more readily controllable.

### The objectives

#### *Equality of focal lengths*

The focal lengths of the two objectives may differ only very slightly. For convenience we shall regard the two objectives as consisting of a single

lens each. When two lenses with different focal lengths have a common object (i.e. common object plane) they must be situated at different points along the axis if they are both to project a sharp image on a common image plane (fig. 6). The object is then projected with different degrees of magnification

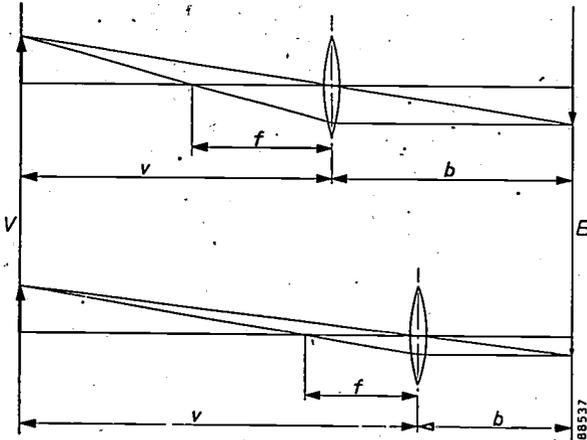


Fig. 6. Two lenses of different focal length, but with a common object plane (V), must be situated at different positions in order that both project a sharp image on a common image plane (B). The magnifications given by the lenses then differ.  $v$  = object distance,  $b$  = image distance,  $f$  = focal length.

(the formula for magnification being  $m = b/v$ , where  $b$  is the image distance and  $v$  the object distance). One consequence of the difference in magnification is that the lines of the odd and even rasters on the film are not equal in length. Thus, if a vertical line in the middle of a film frame is correctly reproduced on the television picture tube, vertical lines at some distance from the middle will show the horizontal image-doubling discussed in the preceding section. Moreover, the two rasters will not be correctly interlaced over the whole of the film frame. If the interlace is correct in the middle of the frame, it will deviate above and below, resulting in vertical image-doubling at the top and bottom of the picture tube.

Since the width of a television picture is greater than its height, and image-doubling increases with the distance from the middle (assuming that the picture is well adjusted in the middle) horizontal image-doubling is the more serious fault. In this connection we have stipulated that at the vertical edges of the picture the reproductions of the same line by the odd and even rasters should not be relatively displaced by more than  $1/4$  of the line spacing. (The line spacing is the vertical distance between two adjacent lines of the odd and even rasters.) This means that the widths of the two rasters on the film may differ by at most half the line spacing on the film

(viz.  $1/4$  at the left and  $1/4$  at the right extremity). The width of a television picture being about  $4/3 \times 625 \approx 800$  line spacings, the magnifications caused by the two objectives should not differ by more than about  $1/1600 \approx 0.06\%$ . Given this condition, a simple calculation shows that the difference between the focal lengths of the optical systems should be less than about  $0.04\%$ .

From the lens formula  $1/b + 1/v = 1/f$ , the magnification formula  $m = b/v$  and the stipulation  $b + v = \text{constant} = c$  (see fig. 6) a relation follows between the focal length  $f$  and the magnification  $m$ :

$$f = c \frac{m}{(1+m)^2}$$

Taking first the logarithm, and then differentiating, we find:

$$\frac{df}{f} = \frac{1-m}{1+m} \frac{dm}{m} \dots \dots \dots (4)$$

We have already seen that the permissible relative variation in magnification is  $0.06\%$ , so that  $dm/m = 0.6 \times 10^{-3}$ . In the present case  $m \approx 1/3$ . By inserting this value and that for  $dm/m$  in (4) we find the permissible relative difference between the focal lengths to be:

$$\frac{df}{f} = 0.4 \times 10^{-3} = 0.04\%$$

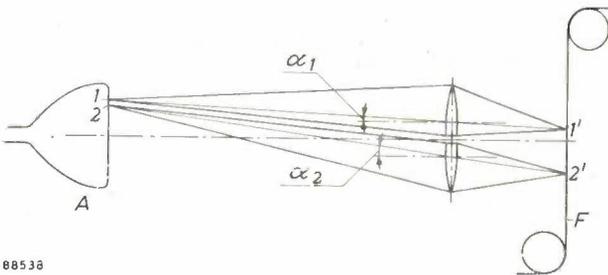
The objectives used are made up of several components. Owing to the tolerances in the spacing between the components, there is often a variation of several per cent in the focal lengths, even in objectives from the same production series. By very slightly modifying the spacing of the components in one of the objectives it is possible nevertheless to make the two focal lengths equal to within the prescribed accuracy of  $0.04\%$ . This alters the lens correction but does not worsen it. The method of testing the equality of the focal lengths, in itself quite a problem, cannot be discussed within the scope of this article.

*Vignetting*

Another requirement to be satisfied by the objectives is that the illumination of the image point should decrease only very slightly with angular deviation from the optical axis, i.e. when the corresponding object point is moved further away from the optical axis. The scanning lines 1 and 2 in fig. 7 will make this clear. On the scanning tube these lines are next to each other. Their active images, however, are produced by different optical systems, i.e. by the upper and lower objectives respectively. The two beams of light make different angles with the optical axes of the objectives and therefore the two images receive different quantities of light.

On the picture tube the lines again appear one above the other, but with alternating luminous intensity. As a result, the picture flickers at a frequency of 25 c/s, particularly at the upper and lower edges, where the differences are most pronounced. In the middle of the picture the odd and even scanning lines receive equal quantities of light.

Since the eye is highly sensitive to flicker of such a low frequency<sup>6)</sup>, the maximum difference between the quantities of light received by the different scanning lines should not amount to more than a few per cent. In every lens the intensity of light in an image point diminishes as the angle  $\alpha$  of the incident light increases, for the light at the image point is proportional to  $\cos^4\alpha$  (the  $\cos^4$  law). If the lenses give rise to vignetting, however, the phenomenon is much more serious. By vignetting of a lens is meant the increasing interception of the light beam, by *multiple* rims or diaphragms, with increasing  $\alpha$ . The objectives used in the present flying-spot scanner have been so designed as to cause no vignetting at the values of  $\alpha$  concerned, thus keeping flicker within permissible bounds.



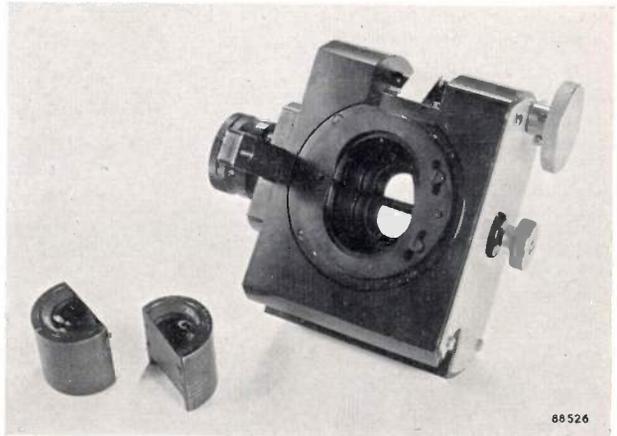
88538

Fig. 7. The light in an image point becomes weaker the further the object point is removed from the optical axis. The image 2' of the raster line 2 projected by the lower objective will therefore receive less light than image 1' of raster line 1 projected by the upper objective. If this difference is not kept within certain limits it becomes noticeable as a 25 c/s flicker in the final picture.

#### Further particulars of the objectives

It is also necessary to ensure that the objectives cause no distortion. Slight distortion would not be so serious if both rasters were distorted in the same way, so that the lines would still interlace properly on the film. This is not the case, however; even very slight distortion causes some loss of interlace.

As discussed, the images projected by the objectives must be about 9.5 mm apart, which means that the distance between the axes of the lenses must be somewhat less than 9.5 mm (fig. 3a). To make



88526

Fig. 8. The lens mount for the double objective. The small central knob on the right is used for changing the spacing between the objectives so as to obtain correct interlacing. Segments have been sawn off both objectives systems as shown.

this possible, it was necessary to saw off segments of the objectives, as illustrated in fig. 8<sup>7)</sup>.

#### The condenser lenses

Behind the film are mounted two condenser lenses which collect all the light passing through the film from the two rasters and concentrate it upon the photo-cathode of a photo-electric cell. Each lens projects the image of its corresponding objective upon the photo-cathode (thus the spot moving over the film does not appear as a sharp image on the photo-cathode). Consider, for example, the light beam from the upper optical system (fig. 3a). Irrespective of the position of the source on the scanning tube, a ray of light passes through each point of the objective (which is the object for its corresponding condenser lens) which is then transmitted to the corresponding image point on the photo-cathode. The light beam is thus continuously spread over the entire image of the objective on the photo-cathode, so that the effects of local variations of sensitivity in the latter are eliminated. Moreover, by correctly adjusting the distance between the axes of the condenser lenses, the images which they project of the corresponding objectives can be made approximately to coincide (fig. 3b). (They can never coincide exactly because of segments having been sawn off the objectives). As a result, the photo-electric cell is equally sensitive for both rasters.

<sup>7)</sup> For the selection of the most suitable make and type of objective, the adjustment of their components and the removal of the segments, we are indebted to the optical department of Philips Research Laboratories, under the direction of P. M. van Alphen.

<sup>6)</sup> J. Haantjes and F. W. de Vrijer, Flicker in television pictures, Philips tech. Rev. 13, 55-60, 1951/52.

**Slip caused by film shrinkage**

The method of film scanning discussed in the foregoing is based on the assumption that the film moves at a constant speed of 25 frames per sec. Where there is shrinkage in a film the speed of transport should be adjusted in order to maintain this value. In fact, however, no provisions are made for matching film speed to shrinkage. The film is driven via a sprocket with four pairs of teeth per picture (each frame having four perforations down each side). The sprocket rotates at a constant speed corresponding to 25 frames per sec which is thus the average speed of the film. The film speed is not, however, entirely constant: if the pitch of a tooth is equal to  $\frac{1}{4}$  of the nominal frame pitch, a certain amount of slip will occur four times per picture if the film is shrunk. The reason why this causes no difficulties becomes clear if we consider rather more closely what the relative positions of the two rasters in the film plane are required to be in order to obtain good interlace.

In the first place the distance between the lower extremities of both rasters in the film plane should be exactly equal to the distance travelled by the

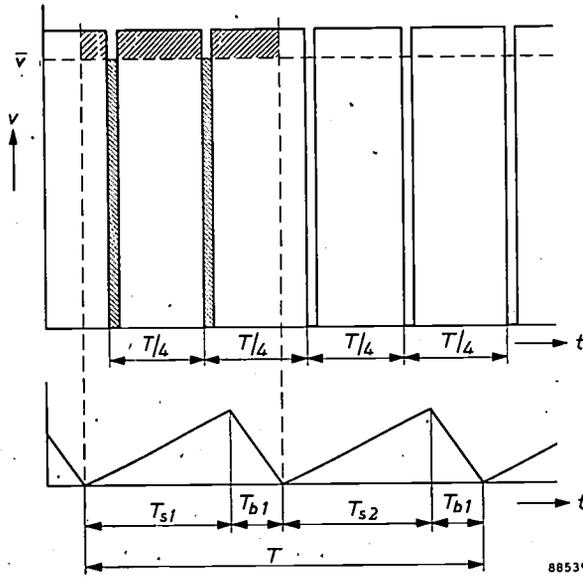


Fig. 9. Film velocity  $v$  as a function of time  $t$  when using a drive sprocket with four pairs of teeth per film frame. Slip occurs four times per picture period, so that the prescribed average speed  $\bar{v}$  of 25 frames per sec is maintained irrespective of shrinkage. It is assumed that the film is stationary during slip. During an interval  $T_{s1} + T_{b1}$  the hatched area above the line  $v = \bar{v}$  is equal to that below the line. The lower figure represents once again the vertical deflection of the spot as a function of time.

film during the period  $T_{s1} + T_{b1}$ . If this is not so, the first lines of the even raster on the film will not fall correctly between the lines of the odd raster. Assuming for the sake of simplicity that the film is

stationary for a moment during each slip, we can plot film speed against time to obtain a curve as shown in fig. 9. The curve shows that during a period  $T_{s1} + T_{b2}$  the film has travelled the same distance as it would have covered if its speed had the proper

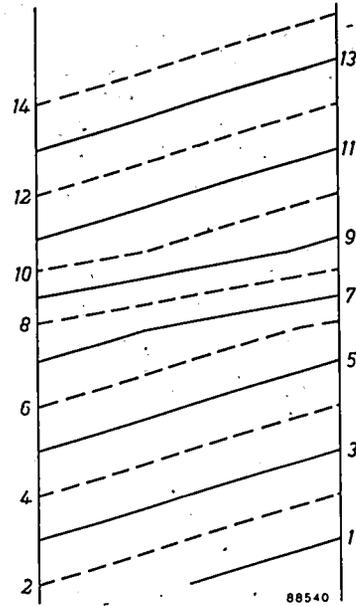


Fig. 10. Distortion of the raster on the film owing to slip. The numbers indicate the sequence of the scanning lines. There is no loss of interlace.

constant value (which is indeed equal to the average speed  $\bar{v}$  in fig. 9). The distance between the two objectives must therefore be adjusted to the same value, independent of whether the film actually has a constant velocity or slips.

In the second place, the interlace — which is now correct at the lower extremities — should remain correct and not be affected by slip. If we have a film with a shrinkage of 1% (the maximum occurring in practice) the film will be stationary during each slip for  $\frac{1}{4}\%$  of the full picture period, during which time  $\frac{1}{4}\%$  of 625 lines, viz.  $1\frac{1}{2}$  lines, will be traced. The vertical velocity of the spot with respect to the film is thus too small during the moments of slip, so that  $1\frac{1}{2}$  lines of, say, the odd raster are traced with too flat a gradient (see fig. 10). Owing to the even number of pairs of sprocket teeth per frame and their uniform spacing,  $1\frac{1}{2}$  lines of the even raster will be traced half a picture period later with the same distortion (dashed lines in fig. 10). Thus, the interlace remains correct throughout. It is clear that this would not be the case if there were, for example, three pairs of sprocket teeth per frame, that is to say an odd number of teeth. The result, with a film having maximum shrinkage (1%), is that the raster on the

picture tube will be slightly distorted over 3 to 4 lines, which is not, however, perceptible. As may be seen from fig. 9, distortion of this nature generally occurs at two positions in the picture. If the film is not stationary during slip, the distortion is spread over the whole picture, but at the same time it becomes less pronounced and thus remains imperceptible in the television picture.

**Summary.** A flying-spot scanner for televising 35 mm film has been developed in the Philips Research Laboratories, Eindhoven. In such scanners the film moves with constant speed and this complicates the problem of interlaced scanning. The problem is solved by using two optical systems. The two objective lenses have to be closely identical and should be as free as possible from distortion and vignetting.

Since there is always some shrinkage in a film and since the drive is effected in the normal way via a sprocket, the film does not in fact travel at a strictly constant speed. A detailed explanation is given of why this is not troublesome.

## ENTROPY IN SCIENCE AND TECHNOLOGY

### IV. ENTROPY AND INFORMATION

by J. D. FAST and F. L. H. M. STUMPERS.

536.75:621.391

*In Volume 16 of this Review one of the authors (J.D.F.) devoted a series of three articles to the application of the concept of entropy to several widely varying problems in science and technology. Some of the subjects discussed were: affinity in chemical reactions, paramagnetism, the attainment of very low temperatures, heat engines, refrigerating machines and heat pumps, radiation, specific heat of solids, vacancies and diffusion in solids, alloys, the elasticity of rubber, and the solution of macro-molecules. This article, which concludes the series, is somewhat different from the previous ones in that its purport is rather negative than positive. In information theory it is possible to arrive at a concept that can be formally defined and manipulated analogously to the concept of entropy in thermodynamics. This is illustrated by various applications. However, the use of the term entropy for this new concept might wrongly suggest that all thermodynamical implications in the term were equally valid here. Some caution is therefore necessary.*

#### Introduction

The concept of information plays an essential part in two recently developed branches of technology, viz. communication theory and cybernetics. The domain of communication theory includes the transmission of messages by means of acoustical, mechanical, electrical or optical signals. Its mathematical foundation has been largely laid down by Shannon<sup>1)</sup>. Even more ambitious in scope is the science of control or *cybernetics*, so termed by Wiener<sup>2)</sup> who has done much fundamental work on this subject. The subtitle of Wiener's book: "Control and communication in the animal and the machine" states succinctly the scope of this new science. More recently<sup>3)</sup> the same author has particularly applied himself to the sociological aspects of his theory. At present a synthesis embracing both communication theory and cyberne-

tics, is developing. In various symposia workers in widely diverging branches of science, such as communication engineers, linguists, psychologists and neurologists have considered to what extent their problems, which are all concerned with the transmission and the use of information, are of a common nature.

In the literature that has sprung up on this subject since about 1945, the term "entropy" is frequently encountered. Mention is made of the entropy of a source of information, the entropy of a language, the entropy of a TV picture, etc. Unfortunately, however, the word entropy is often used here in a sense that bears but little relation to the well-established thermodynamical-statistical entropy, discussed in the preceding three articles of this series<sup>4)</sup>. We shall return to this after having first said something more about communication, cybernetics and information.

<sup>1)</sup> C. E. Shannon, A mathematical theory of communication, Bell System tech. J. 27, 379-423 and 623-656, 1948.

<sup>2)</sup> N. Wiener, Cybernetics, Hermann at Cie., Paris 1947, John Wiley, New York 1948.

<sup>3)</sup> N. Wiener, The human use of human beings: Cybernetics and society, Houghton Mifflin, Boston 1950.

<sup>4)</sup> J. D. Fast, Entropy in science and technology: I. The concept of entropy, Philips tech. Rev. 16, 258-269, 1954/55; II. Examples and applications, *ibid.* pp. 298-308, III. Examples and applications (cont.), *ibid.* pp. 312-332.

picture tube will be slightly distorted over 3 to 4 lines, which is not, however, perceptible. As may be seen from fig. 9, distortion of this nature generally occurs at two positions in the picture. If the film is not stationary during slip, the distortion is spread over the whole picture, but at the same time it becomes less pronounced and thus remains imperceptible in the television picture.

**Summary.** A flying-spot scanner for televising 35 mm film has been developed in the Philips Research Laboratories, Eindhoven. In such scanners the film moves with constant speed and this complicates the problem of interlaced scanning. The problem is solved by using two optical systems. The two objective lenses have to be closely identical and should be as free as possible from distortion and vignetting.

Since there is always some shrinkage in a film and since the drive is effected in the normal way via a sprocket, the film does not in fact travel at a strictly constant speed. A detailed explanation is given of why this is not troublesome.

## ENTROPY IN SCIENCE AND TECHNOLOGY

### IV. ENTROPY AND INFORMATION

by J. D. FAST and F. L. H. M. STUMPERS.

536.75:621.391

*In Volume 16 of this Review one of the authors (J.D.F.) devoted a series of three articles to the application of the concept of entropy to several widely varying problems in science and technology. Some of the subjects discussed were: affinity in chemical reactions, paramagnetism, the attainment of very low temperatures, heat engines, refrigerating machines and heat pumps, radiation, specific heat of solids, vacancies and diffusion in solids, alloys, the elasticity of rubber, and the solution of macro-molecules. This article, which concludes the series, is somewhat different from the previous ones in that its purport is rather negative than positive. In information theory it is possible to arrive at a concept that can be formally defined and manipulated analogously to the concept of entropy in thermodynamics. This is illustrated by various applications. However, the use of the term entropy for this new concept might wrongly suggest that all thermodynamical implications in the term were equally valid here. Some caution is therefore necessary.*

#### Introduction

The concept of information plays an essential part in two recently developed branches of technology, viz. communication theory and cybernetics. The domain of communication theory includes the transmission of messages by means of acoustical, mechanical, electrical or optical signals. Its mathematical foundation has been largely laid down by Shannon<sup>1)</sup>. Even more ambitious in scope is the science of control or *cybernetics*, so termed by Wiener<sup>2)</sup> who has done much fundamental work on this subject. The subtitle of Wiener's book: "Control and communication in the animal and the machine" states succinctly the scope of this new science. More recently<sup>3)</sup> the same author has particularly applied himself to the sociological aspects of his theory. At present a synthesis embracing both communication theory and cyberne-

tics, is developing. In various symposia workers in widely diverging branches of science, such as communication engineers, linguists, psychologists and neurologists have considered to what extent their problems, which are all concerned with the transmission and the use of information, are of a common nature.

In the literature that has sprung up on this subject since about 1945, the term "entropy" is frequently encountered. Mention is made of the entropy of a source of information, the entropy of a language, the entropy of a TV picture, etc. Unfortunately, however, the word entropy is often used here in a sense that bears but little relation to the well-established thermodynamical-statistical entropy, discussed in the preceding three articles of this series<sup>4)</sup>. We shall return to this after having first said something more about communication, cybernetics and information.

<sup>1)</sup> C. E. Shannon, A mathematical theory of communication, Bell System tech. J. 27, 379-423 and 623-656, 1948.

<sup>2)</sup> N. Wiener, Cybernetics, Hermann at Cie., Paris 1947, John Wiley, New York 1948.

<sup>3)</sup> N. Wiener, The human use of human beings: Cybernetics and society, Houghton Mifflin, Boston 1950.

<sup>4)</sup> J. D. Fast, Entropy in science and technology: I. The concept of entropy, Philips tech. Rev. 16, 258-269, 1954/55; II. Examples and applications, *ibid.* pp. 298-308, III. Examples and applications (cont.), *ibid.* pp. 312-332.

## Communication

It is self-evident that communication plays a predominant role in modern civilization. No society however, primitive, is conceivable without communication — primarily in the form of a language of gestures and spoken words. In modern society this is augmented by the written and printed word, photography, telegraphy, telephony, teleprinter, radio, the cinema and television. The word communication means, quite generally, any transmission of information from one place to another. The entities and the media that participate in this constitute together a communication system. In its simplest form such a system (*fig. 1*) consists of a source of information plus a transmitter at the one end, and a receiver plus a recipient at the other end, with a communication channel linking the two.

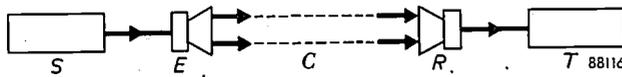


Fig. 1. Schematic representation of a communication system. S source of information (selector), E transmitter + coding device, C communication channel, R receiver + decoding device, T recipient.

If the communication consists of something spoken by a person *A* to a person *B*, the brain and the organs of speech of *A* constitute the source and the transmitter, whilst the ear and the brain of *B* are the receiver and the recipient. The communication channel in this case is the air.

The source of information preselects a message, i.e. it chooses a specific sequence of the words, letters or figures in which the message is embodied. From this the receiver tries to reconstruct the original message, after which it is passed along to the recipient. In the case of the spoken word the signal consists of sound waves in the air, in the case of line telephony of variations in the electric current in a wire. The transmitter in the latter case is a microphone, which converts the varying sound pressure of the voice into a varying electric current; the telephone is part of the receiver. The sound waves in the air and the varying electric current in the wire correspond to the message to be transmitted according to a certain "code".

The application of carrier-wave telephony<sup>5)</sup> makes it possible to transmit several telephone conversations over a single line, i.e. over one conductor pair. The low-frequency electric vibrations derived from corresponding acoustic vibrations

from speech, are here superimposed on the high-frequency vibrations of the carrier wave. When carriers of different frequencies are separately modulated in this way with audio-frequency speech, several conversations can be transmitted at the same time. The substantially increased efficiency of information-transmission thus brought about gives rise to the question: what is the maximum amount of information per unit time that can be transmitted over a single telephone line? Before we can answer this question we need a measure to express "amount of information". The fact that there is a finite limit to the capacity of a communication channel to carry information follows from the inevitable occurrence of thermal agitation noise (see I) in every communications system. This blurs the details of the transmitted symbols, thus limiting their intelligibility. Only after we have introduced a measure for information can we deal with the above-mentioned question.

## Cybernetics

Cybernetics is primarily concerned with self-regulating systems and processes. A very simple example of a self-regulating system is a thermostat, with which a space is automatically kept at a nearly constant temperature. This form of control can be effected, for example, with a special type of thermometer in which the mercury closes a contact as soon as the temperature exceeds a predetermined value. The electrical circuit is arranged so that a relay is tripped which reduces the supply of heat. When subsequently the temperature drops below the desired value, the thermometer contact is broken and the relay automatically causes an increased supply of heat. The essence of the matter is that here we are not concerned with a one way sequence of cause and effect, as with a room with a normal thermometer, where the position of the mercury in the thermometer depends simply on the temperature of the room; instead, we have here a two-way sequence of cause and effect, whereby the temperature of the room is also influenced by the position of the mercury.

We may put it that the above automatic control is due to the fact that the information regarding the temperature is fed back from the thermometer to the source of heat. This *feedback* is the common characteristic of the systems with which cybernetics is concerned. The information regarding a deviation from the desired condition is fed back to the control mechanism, which solely on the strength of this information can fulfill its controlling action.

<sup>5)</sup> See e.g. H. N. Hansen and H. Feiner, Coaxial cable as a transmission medium for carrier telephony, (Philips tech. Rev. 14; 141-150, 1952/53).

A more elaborate example of automatic control by information-feedback is the automatic aiming of an anti-aircraft gun. A radar installation supplies data regarding the position of an aircraft to an electronic device, which, via servomotors, moves the gun into the appropriate direction. In order to function properly and to give the right instructions to the servomotors the electronic device must not only know the position of the aircraft, but also that of the gun. Of the latter it is automatically informed by electric signals transmitted by a device connected to the gun. These signals thus again provide the necessary feedback of information.

Control instruments of the latter kind are further characterized by the fact that they predict the probable future position of the aircraft (or other variable) from the data so far received. The mathematical theory of these prediction problems has been examined by Wiener and Kolmogoroff.

The ambitious scope of the cyberneticists will be better appreciated when one reflects on the many automatic processes that are continuously occurring in living creatures e.g. in the human body: its breathing, its heartbeat, its reaction to stimuli, etc. Many of these processes are involuntary, being controlled by reflexes, which are formally comparable to the above-mentioned mechanisms employing the feedback principle. An example is the flexor-reflex, which occurs for instance if one inadvertently touches a hot object with the hand: Along the sensory nerves a signal in the form of series of electromechanical impulses of short duration is passed to the spinal cord. Here the information is passed on to certain motor neurones, along which commands are sent to the flexor muscles of the arm in question, resulting in the withdrawal of the hand from the hot object. Regulation by means of hormones (organic substances secreted by certain glands in the body) may likewise be considered as a feedback mechanism. A well-known hormone is insulin, which is produced in the pancreas and plays an important part in regulating the metabolism of carbohydrates. We are probably justified in saying that all vegetative functions in the body, e.g. the control of the composition, the pressure and the pH of the blood, the osmotic pressure, the body temperature, and the composition of the gas in the lungs, are based on a form of feedback, either hormonal or nervous in character. This control has the task of providing the tissues with a constant environment, which is indispensable to the proper functioning of the organism. The name "homeostasis" has been given by Cannon <sup>6)</sup> to this system of regulation.

<sup>6)</sup> W. B. Cannon, *The wisdom of the body*, Paul, Trench, Trubner & Co., London 1939.

With any automatic control by feedback alone we have to put up with minor deviations from the desired value, since it is precisely these deviations that are employed to execute the necessary corrections. We are thus always concerned with oscillations around the desired value and it is our aim to reduce these to a minimum. Some cyberneticists believe that it will eventually become possible to mitigate the recurring economic crises of our society by the application of the methods of feedback engineering. They proceed from the premise that these economic fluctuations are based on a feedback of information in the sense that economic activity is determined by the rate of investment, but that the latter, in connection with the profits to be expected, is in its turn dependent upon economic activity <sup>7)</sup>. They hold that by taking certain timely action it should in principle be possible to put a curb on economic fluctuations.

A field of great importance is opened up by the application of the principles of cybernetics to the development of "automatic factories", that is the gradual but increasing change to almost wholly automatic plants and machines.

With this "automation" the human element is entirely or partially replaced by robots that act in a controlling and regulating capacity upon the manufacturing process. Like the thermometer in the thermostat, the robot performs both the measuring and the regulating functions. It may be given the power of human muscles, or more than human power by amplification, and may even be given, in the form of an electronic computer, some of the elementary faculties of the human mind. Modern electronic computers are capable of completely taking over certain secondary functions of the human brain and within these restricted fields of action often improve on the brain with regard to speed and accuracy.

#### The concept of information

In daily life we speak of receiving information when we learn something that was yet unknown to us: Information removes uncertainty, or rather, *may* remove uncertainty. For the communications engineer only the latter is of importance. He is interested only in how much information can be contained in a message, and how much difficulty is involved in transmitting it. Its significance to the receiver or its intrinsic value do not concern him.

<sup>7)</sup> A. Tustin, *The mechanism of economic systems: An approach to the problem of economic stabilisation from the point of view of control-system engineering*, Heinemann, London 1953.

For the communications engineer all messages of equal length (i.e. symbol content) are equivalent.

For the person who wishes to transmit a message the choice of the symbols is of considerable importance. If the message to be transmitted consists, e.g. of a series of 1000 symbols, chosen from an alphabet of 32 symbols (26 letters + some punctuation symbols) then the message will be one out of a total of  $32^{1000}$  possible messages. If the message were to contain only one symbol, there would be 32 possibilities; if the message were 2 symbols long, then each of the 32 symbols might be followed by any of 32 symbols, so that there would be  $32^2$  possibilities, etc. In normal practice, as regards conversation, telephony and telegraphy, one is restricted to those series of symbols that constitute an intelligible message in one of the known languages or in an accepted code. For a message of 1000 symbols one then still has the choice of a multitude of possibilities, but considerably less than  $32^{1000}$ .

Let us, for the time being, consider the symbols as independent of each other and occurring with equal probability; then the general expression for a number  $q$  of possible messages of length  $N$  symbols, selected from  $m$  different symbols is  $q = m^N$ . Let another message have a length of  $M$  symbols, then the number of possible selections is  $q' = m^M$ . If the two messages are transmitted in direct succession, and are considered as a single message, then a choice has been made out of  $Q = m^N \times m^M$  possible messages. On the other hand it is only logical to claim that the total service rendered by the telegraph office is the sum of the two separate services. These services consist of the transmission of information. We must, therefore, find a definition of the concept "quantity of information"  $I$  as a function of the number of possible messages, in such a form that:

$$I(m^N) + I(m^M) = I(m^{N+M}).$$

To express the additive character of the function, the quantity of information in a message of a given length is defined as the logarithm of the number of possible messages of that length. For the case considered here we can put:

$$\begin{array}{l} \text{First quantity of information} : \log q = N \log m \\ \text{Second quantity of information} : \log q' = M \log m \\ \hline \text{Total quantity of information} : \log Q = (N+M) \log m. \end{array}$$

Any separate symbol from the series of  $m$  equally probable and mutually independent symbols thus represents a quantity of information

$$i = \log m = -\log p, \quad (\text{IV}, 1)$$

where  $p = 1/m$  is the *a priori* probability of the symbols. The choice of the base of the logarithms is quite arbitrary since this merely determines the magnitude of the unit. If a choice of two equally probable possibilities is transmitted, then the quantity of information supplied is, according to (IV, 1),  $i = \log 2$ . This is the simplest situation conceivable, and it is therefore reasonable to choose this quantity of information as the unit:  $\log 2$  assumes the value unity if 2 is chosen as the base of the logarithms. This has in fact been generally adopted in communication theory. This unit of information is called a "bit" (short for "binary digit"). Formula (IV, 1) appropriately allots a higher numerical value to information according as the freedom of choice is greater, i.e. according as the result is more uncertain (less probable).

It will be clear from the foregoing that the use of a logarithmic function for the definition (IV, 1) is analogous to the logarithmic definition (I, 12) of statistical-thermodynamical entropy: both information content and entropy are additive, while the probabilities of independent events are multiplicative in both cases (see I, pp. 265-266).

#### The "entropy" of a series of events

A single symbol from the language containing  $32 = 2^5$  equally probable and mutually independent symbols, supplies, according to formula (IV, 1), 5 bits of information. The symbols of a language are, in fact, not equally probable; moreover they cannot always occur independently. In order to estimate of the actual information content per symbol, we shall first of all get rid of the restriction that the symbols are equally probable. Later we shall also drop the restriction that they are independent.

Let us begin by considering an imaginary language in which the symbols occur with unequal probability and are independent of each other. For any given language the probabilities are immediately apparent from the frequencies with which the symbols occur in a number of random texts in that language. The frequency of letters in the English language is given below in Table I<sup>8)</sup>.

It is found that letter frequencies are only slightly influenced by the subject and the author of the text.

The quantity of information conveyed by a single letter we shall define by analogy with formula

<sup>8)</sup> Taken from Fletcher Pratt, *Secret and Urgent*, Blue Ribbon Books, New York 1942.

Table I. Probabilities of occurrence of the various letters in the English language.

Letter	Probability	Letter	Probability
e	0.131	m	0.025
t	0.105	u	0.025
a	0.086	g	0.020
o	0.080	y	0.020
n	0.071	p	0.020
r	0.068	w	0.015
i	0.063	b	0.014
s	0.061	v	0.0092
h	0.053	k	0.0042
d	0.038	x	0.0017
l	0.034	j	0.0013
f	0.029	q	0.0012
c	0.028	z	0.00077

(IV, 1) as  $-\log_2 p$ ,  $p$  representing the relative frequency with which this letter occurs. The question now arises how much information will be conveyed on the average by a symbol from an English text. There is a chance  $p_e$  that it will be an  $e$ , in which case a quantity of information  $-\log_2 p_e$  will be received. There is further a chance  $p_t$  that it will be a  $t$ , in which case a quantity of information  $-\log_2 p_t$  will be received, etc. The average amount of information per symbol received is therefore

$$H = -\log_2 p_v = -\sum_v p_v \log_2 p_v \text{ bits per symbol.} \quad (\text{IV, 2})$$

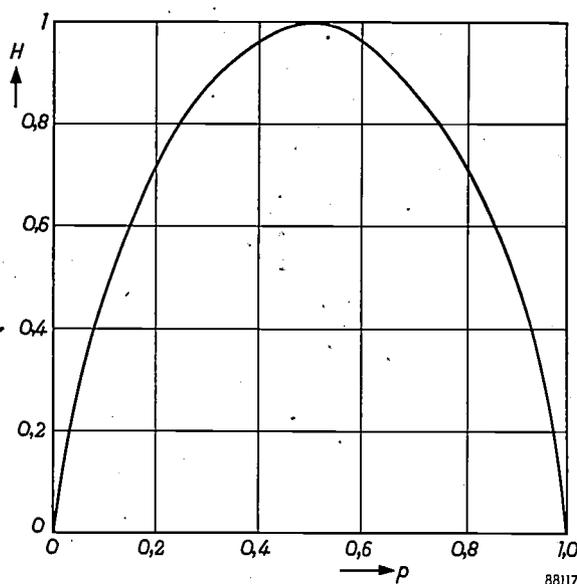
With the exception of a constant we have met a formula of the same form in (I, 16), where it served as a definition of *entropy* in statistical thermodynamics. Following Shannon, the above-defined quantity  $H$  has also been designated "entropy". In the opinion of the present authors, this is a regrettable practice, not to be recommended. Instead, some such term as *information-content per symbol* is to be preferred, in view of the fact that the relationship between this information-entropy and the thermodynamical-statistical entropy is only of a formal mathematical nature, due to the common statistical background. The thermodynamical concept of entropy, however, also has physical implications: it is directly connected with the concept of temperature and with the quantity of heat that can be exchanged in the course of a process between a system and its surroundings (see I).

The definition of the (average) quantity of information per symbol, or entropy, is generally valid for series of events that can each occur with a certain probability. If only two events are possible — with probabilities  $p$  and  $(1-p)$  — then the

average amount of information per event is given by

$$H = -p \log_2 p - (1-p) \log_2 (1-p). \quad (\text{IV, 3})$$

In fig. 2,  $H$  is plotted as a function of  $p$ . We see that  $H$  reaches a maximum ( $= \log_2 2$  or 1 bit) when the two possibilities are equally probable, i.e. when  $p = 1 - p = 0.5$ , in other words, when the *a priori* uncertainty with regard to the result is greatest. In the extreme cases,  $p = 1$  (certainty) and  $p = 0$  (impossibility), formula (IV, 3) gives  $H = 0$ , since then there exists no uncertainty whatsoever. Returning from formula (IV, 3) to the general formula (IV, 2), we observe that in this case too

Fig. 2. The function  $H = -p \log_2 p - (1-p) \log_2 (1-p)$ .

the quantity of information per event (i.e.  $H$ ) is a maximum if all probabilities are equally great, which means that there is a maximum freedom of choice, that is to say, there is the greatest uncertainty with regard to the outcome.

In connection with the above considerations, J. F. Schouten<sup>9)</sup> has pointed out that the activity or "manipulation" leading to the information should be distinguished from the information itself, the latter being merely passive. Schouten expresses the quantity of manipulation in *bics* (binary choice); it is equal to the maximum quantity of information (in bits) that one may expect from a process. This distinction between the "active" unit bic and the "passive" unit bit has not been generally recognized.

As an illustration we may mention the well-known radio parlour-game "Twenty Questions" in which the team has to find the name of an object

<sup>9)</sup> J. F. Schouten, Proceedings of the Symposium on Information Theory, London 1950, page 195.

by putting questions that can be answered only by yes or no. The best policy for the team is to put questions that have an equal chance of being answered by yes or by no. A simple example to illustrate this principle is the following. If an object is to be located on a chessboard ( $64 = 2^6$  squares), then the quickest procedure is to divide the board each time into halves (lower half, upper half, left half, and right half, etc.). In this way only 6 questions are necessary. Somewhat more complicated is the following problem. We have a pair of scales (without weights) and twelve coins of identical appearance. One of the coins is counterfeit. The question is how many weighings are at least required to decide which coin is counterfeit and whether this coin is too light or too heavy. We must, therefore, locate one coin out of twelve and furthermore make a choice of two possibilities. This requires  $\log_2 12 + \log_2 2 = 4.585$  bits of information. One weighing may have 3 results: equilibrium, the scale tips to the left, the scale tips to the right. We can, therefore, expect  $\log_2 3 = 1.585$  bits of information at the most from each weighing. For this reason it may be expected that 3 weighings will be sufficient provided that they are carried out properly. If we arrange our experiments in such a way that the 3 possible results of every weighing are about equally probable, this can indeed be realized, though not without some trouble.

The total amount of information required ( $\log_2 24$  bits) is less than the maximum amount of information that can be derived from three weighings ( $3 \times \log_2 3 = \log_2 27$  bits). Consequently to solve the problem in three weighings a number of slightly different strategies may be possible. An example of a good strategy is the following.

Divide the twelve coins into three groups of 4 and weigh two groups against each other. If the scale dips, then we know that one of these groups of 4 may contain a coin that is too light or the other group of 4 may contain a coin that is too heavy. For the second weighing we place on each scale pan two of the coins that are possibly too light together with one of the coins that are possibly too heavy. If there is equilibrium then we know that one of the 2 remaining coins is counterfeit and too heavy. A single weighing will then show which of them is counterfeit. If there is no equilibrium then we weigh the two possibly-too-light coins that were on the scale-pan that went up, against each other. If with this third weighing the scale dips, then the coin on the scale-pan that goes up is counterfeit and too light. If there is equilibrium however then the coin that was possibly too heavy on the scale-pan that went down in the second weighing is the culprit.

There remains the possibility that in the first weighing the scales were balanced. We then take three of the four remaining coins and one of the 8 good coins and place 2 of these four coins on each scale. If there is equilibrium then the remaining twelfth coin is counterfeit and a single weighing will establish whether it is too light or too heavy. If there is no equilibrium, then we weigh the two coins on the pan not containing the good

coin against each other. The result of this third weighing is then decisive. It is instructive to evaluate the information obtained from these experiments by means of formula (IV, 2) for each weighing. A similar puzzle concerns 13 coins of identical appearance, one of which may be counterfeit. Here the problem is to find the counterfeit coin (if there at all) and to decide whether it is too light or too heavy in three weighings. A 14th, good standard coin is provided. In this problem both the required and the maximum obtainable quantity of information are  $\log_2 27$  bits, and there exists only one good strategy.

### The "entropy" of a language

If we evaluate the average quantity of information per letter of the English language from the probabilities given in Table I, then we obtain 4.16 bits/letter. If the 26 letters had had equal probabilities, then  $H$  would have been  $\log_2 26 = 4.70$  bits/letter. The quantity of information per letter has therefore been reduced<sup>10)</sup>, even for an independent choice. Moreover, the probability of a letter is also somewhat dependent on the letters earlier transmitted or written down. We shall illustrate this with 2 examples: 1) The probability of the letter  $u$  is not particularly great, but the probability that a  $u$  follows a  $q$  approaches unity. 2) If we have received, as part of a message, the following symbols: Arrived Saturday mor. . . ." then it is highly probable that the next letters will be "ning", although "ose", "e", "eover" or "occo" are not inconceivable. Many other examples of correlation could be given. They all show that the number of different meaningful messages which are possible in a message of a given length is reduced by the presence of correlation.

An artifice to compute roughly the quantity of information per symbol is the following: let  $p_\nu$  in formula (IV, 2) represent the probability of the occurrence of super-symbols each of length  $A$  symbols. According as we make  $A$  larger, the inter-correlations between the super-symbols will become smaller. The super-symbols, if sufficiently long, can therefore be chosen nearly independently of each other. In a very long message of  $N$  super-symbols there will be on the average  $p_1 N$  specimens of the first super-symbol,  $p_2 N$  specimens of the second super-symbol, etc. For the information per super-symbol we then obtain, according to (IV, 2):

$$-\sum_{\nu} p_{\nu}(A) \log p_{\nu}(A).$$

<sup>10)</sup> When determining the probability per symbol we should actually also take the spacings and punctuations into account. This has not been done here, in order to make use of existing tables.

Since each super-symbol comprises  $A$  symbols, the average information per symbol is represented by:

$$H = -\lim_{A \rightarrow \infty} \frac{1}{A} \sum_p p_p(A) \log p_p(A) \text{ bits per symbol.} \quad \dots \text{(IV, 4)}$$

An increasingly closer approximation to  $H$  can be reached by choosing successively 1, 2, 3, ..... for  $A$ , i.e. by determining the frequencies with which single symbols, symbol pairs, symbol triples, etc. occur in the language. Another estimate of the quantity of information per symbol can be arrived at by establishing experimentally how many times a person has to guess before he has found the next letter in a given text, and then the next letter, etc. There have not yet been enough experiments of this kind made to establish accurately the actual information content per symbol of a language, but it has been found that for the modern European languages the information content is around 1.5 bits per symbol.

The part played by the statistical structure in a language has been aptly demonstrated by Shannon<sup>11)</sup>. He first considers a sequence of symbols constructed from a 27-letter alphabet (26 letters and a space) each symbol being chosen quite randomly (i.e. with equal probability) and independently. This can be regarded as a zero-order approximation to any language using the roman alphabet. The first-order approximation to (say) English is then obtained by choosing successive letters independently but with each letter and the word spacing occurring with the natural frequency that they have in English (see Table I). Thus  $e$  is chosen with probability 0.13 (roughly its relative frequency in English) and  $w$  with probability 0.015. For the second order approximation, pair-structure<sup>11)</sup> is introduced: after a letter is chosen, the next one is chosen in accordance with the natural frequencies of the various pairs (e.g.  $th$ ,  $ed$ , are frequent pairs in English). In the third-order approximation, triple-structure is introduced: each letter is then chosen with probabilities which depend on the preceding two letters.

<sup>11)</sup> Pair and triple frequencies are also given in the book by Fletcher Pratt quoted in footnote<sup>8)</sup>. Word frequencies are tabulated in Relative frequency of English speech sounds, by G. Dewey, Harvard University Press 1923. Shannon<sup>1)</sup> also mentions a simple method for constructing these examples without the use of such tables: To construct a second order letter approximation, for example, open a book at random and select a letter at random on the page. This letter is recorded. The book is then opened to another page and one reads until this letter is encountered. The succeeding letter is then recorded. Turning to another page this second letter is searched for and the succeeding letter is recorded; and so on. This process was actually used by Shannon for constructing the examples (c) and (d) and for the word approximations (e) and (f).

Proceeding in this way we can construct better and better approximations to English:

- a) Zero-order approximation (symbols independent and all equally probable).  
*xfoml rxkhrjffjuj zlpwcfwkcyj*  
*ffjevvcqsgxyd qpaamkbzaacibzlhjgd*
- b) First-order approximation (symbols independent but with frequencies as in English).  
*ocro hli rgwr nmiehwis eu ll nbnesebya th eei*  
*alhenhtpa oobttva nah brl*
- c) Second-order approximation (pair-structure as in English)  
*on ie antsoutinys are t inctore st be s deamy*  
*achin d ilonasive tucoowe at teasonare fuso*  
*tizin andy tobe seace ctisbe*
- d) Third-order approximation (triple-structure as in English).  
*in no ist lat whey cratict froure birs grocid*  
*pondenome of demonstures of the reptagin is*  
*regoactiona of cre*

In the last example (3<sup>rd</sup>-order approximation) some words and parts of words are beginning to be recognizable. Further approximations become more and more time-consuming and are therefore not attempted.

An analogous synthesis, and one that sooner resembles the language, may be achieved by considering whole words instead of letters:

- e) First-order word approximation (words chosen independently but with their natural frequencies).  
*representing and speedily is an good apt or*  
*come can different natural here he the a in*  
*came the to of to expert gray some to furnishes*  
*the line message had be these.*
- f) Second-order word approximation (the transition probabilities of the words, i.e. their natural pair frequency is taken into account).  
*the head and in frontal attack on an english*  
*writer that the character of this point is there-*  
*fore another method for the letters that the*  
*time of who ever told the problem for an un-*  
*expected.*

Again the resemblance to normal English increases noticeably at each stage.

All these examples clearly demonstrate that the more we take the structure of the language into account the more the choice is restricted. The information content per symbol is consequently smaller than it would be with a free, independent choice. This difference between the maximum information content per symbol and the actual information content is called the *redundancy*  $R$ :

$$R = H_{\max} - H. \quad \dots \text{(IV, 5)}$$

The *relative redundancy* is also often considered; this is the ratio of the redundancy to the maximum information content:

$$r = \frac{H_{\max} - H}{H_{\max}} \dots \dots \dots \text{(IV, 6)}$$

For languages such as English, French, German, Dutch the relative redundancy is 60-70%. (At the beginning of this section, p. 206,  $H_{\max}$  is evaluated at 4.70 for the roman alphabet, whilst on the next page,  $H$  is given as 1.5 for the modern European languages.) For the telegraphic transmission of information in one of the above languages it would in principle be possible to save 70% of the time by converting the messages into such a code as to reduce the redundancy almost to zero. This obviously requires the coding *en bloc* of long series of symbols which have hardly any mutual correlation left. The use of such a complicated code would in its turn cause a certain delay. This might be acceptable if it were not for the fact that a finite redundancy has some advantages of its own. A finite redundancy makes it possible to reconstruct a message mutilated by interference or noise at the receiving end. If we receive a telegram: "A drive next suntay" then we assume immediately that the meaning was "arrive next sunday". In a code with maximum information content per symbol, however, any alteration will completely alter the significance of the message and it is by definition no longer possible to recover the original meaning from the context. Since distortion during transmission is inevitable, irrespective of the system of communication, (noise with electric transmission and ambient noise with the spoken word), a language should always have a certain redundancy. The fact that the redundancies of the known languages, in spite of (or because of?) their very long history, vary only slightly gives rise to the supposition that this "general" redundancy is not far removed from the optimum value<sup>12)</sup>. A slight reduction of the redundancy, however, is possible without harm being done. This is demonstrated by certain of the simplifications adopted in the American version of the English language.

#### The "entropy" of a television signal

The general observations on the "entropy" of a series of events are applicable to every form of communication. We shall deal here briefly with their bearing on television. Suppose that the separate

elements of an ordinary black-and-white television-picture can assume a number  $n$  of different brightness levels. Let us again indicate the probabilities of the occurrence of these levels by  $p_v$ , then the "entropy" of every television picture could be expressed by

$$H = - \sum_{v=1}^n p_v \log p_v \text{ bits per picture element,}$$

if the brightness levels of the separate picture elements were completely independent of each other. (Here, too, it would be better not to speak of entropy but of the quantity of information per picture element.) The brightness levels of the picture elements, however, show a definite correlation with regard to both place and time. Most picture elements in an ordinary picture vary only slightly in brightness from adjacent elements. We would furthermore not get an impression of continuous motion if the majority of the picture elements did not vary gradually in brightness with time. Like a language, the television picture shows a considerable redundancy, and similarly, a picture that only possessed the statistically correct brightness distribution would never approximate to a normal picture. The efficiency of the transmission of information from the television transmitter to the receiver is decidedly impaired by this considerable redundancy. This is manifested by the necessity of a very large bandwidth for the present television systems. Recently some interesting suggestions were made as to how, with the aid of a suitable system of coding, the redundancy and hence the bandwidth of the signal to be transmitted might be reduced<sup>13)</sup>. Information theory defines the limits of what is attainable, but the future must decide how closely our technology can approach them.

#### Capacity of a communication channel with noise

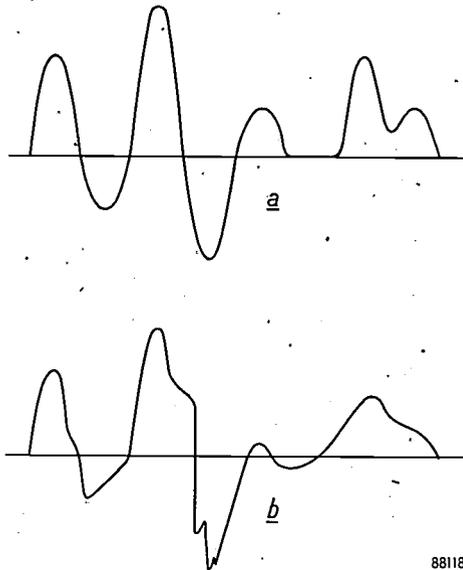
In the foregoing we have mentioned in passing the possible interference with and mutilation of the information content of a message during transmission. As a rule the signals comprising the message in some form of code reach their destination via a transmission path (telephone line, radio link) in which interference occurs. The main source of interference, which is universally present, is random noise. This is caused by the thermal agitation in resistors and valves of transmitter, receiver and amplifying stations, and by many other random fluctuations on the transmission path. Owing to the interfering influence of the noise, we cannot be certain whether the signal received is the one that

<sup>12)</sup> Exhaustive studies on this and other structural phenomena of the language have been made by B. Mandelbrot. See e.g. Structure formelle des textes et communication, Word 10, 1-17, 1945.

<sup>13)</sup> E. C. Cherry and G. G. Gouriet, Proc. Instn. Electr. Engrs 100, 9-18, 1953.

was actually transmitted. The information received is consequently less than it would be if no noise were present. Let us consider a case in which there are a total of  $n$  symbols that can all be transmitted with equal *a priori* probability. The reception of each symbol would then supply, in the absence of noise,  $\log_2 n$  bits of information. If, however, the presence of noise has the effect that reception of a given symbol only means that we have a choice of three equally probable "adjacent" symbols, then we lack  $\log_2 3$  bits of information, and we have thus received only  $\log_2 n - \log_2 3$  bits.

A similar influence of the noise on a communication system, can be imagined as follows: the information contained in a message is coded at the transmitter end in the form of short, similar electric pulses, which are transmitted in rapid succession (fig. 3). Each letter then corresponds,



88118

Fig. 3. Transmission of information by series of sinusoidal pulses. The information is contained in the amplitude of the pulses. At the receiving end (b) the pulses are received in a mutilated condition, so that an unequivocal recognition of the original symbol at the transmitting end (a) is not possible.

say, to a specific pulse height. The effect of noise will now be that the pulses arrive at the receiving end in a mutilated condition, so that there will be some degree of uncertainty regarding the original height of each pulse. According as the influence of the noise is greater compared with the minimum difference in level between two pulses representing different symbols, so the uncertainty regarding each received signal will be greater and the information obtained correspondingly less.

It is possible to formulate these thoughts in a somewhat stricter mathematical form. In the following, some general and well-known formulae are derived concerning the maximum quantity of information

per symbol that can be obtained at the receiving end of a communication system with noise, and concerning the minimum power that is required for the transmission of information. Remarkably enough, with this latter problem, the thermodynamical entropy re-enters the picture.

Let the probability that a given symbol is transmitted be  $p_\nu$ . In the case of a correct transmission of mutually independent symbols, the average information content per symbol at the receiving end would amount to

$$H = - \sum_\nu p_\nu \log p_\nu \dots \dots (IV, 2)$$

This and all subsequent summations have to be made over the whole range of available symbols.

Let  $p_{\mu\nu}$  be the chance that a transmitted symbol  $\mu$  is detected at the receiving end as the symbol  $\nu$ .  $p_{\mu\nu}$  may be called a *transition probability*. In a noise-free channel

$$p_{\mu\nu} = \begin{cases} 1 & \text{if } \mu = \nu. \\ 0 & \text{if } \mu \neq \nu. \end{cases} \dots \dots (IV, 7)$$

If a random, long series of symbols is transmitted, then the nett chance that each received symbol will be  $\nu$  is

$$q_\nu = \sum_\mu p_{\mu\nu} p_\mu \dots \dots (IV, 8)$$

and the chance that any received symbol  $\nu$  is actually the result of the transmission of a symbol  $\lambda$  is

$$P_{\lambda\nu} = \frac{p_\lambda p_{\lambda\nu}}{q_\nu} = \frac{p_\lambda p_{\lambda\nu}}{\sum_\mu p_\mu p_{\mu\nu}} \dots \dots (IV, 9)$$

The reception of symbol  $\nu$  leaves, according to (IV, 9), an additional chance for each symbol that it was *actually* transmitted as such. The uncertainty as to whether the received symbol is the same as the transmitted symbol is therefore

$$h_\nu = - \sum_\lambda P_{\lambda\nu} \log P_{\lambda\nu} \dots \dots (IV, 10)$$

The *average uncertainty* after reception of any symbol is then:

$$\bar{h} = \sum_\nu q_\nu h_\nu = - \sum_\nu \sum_\lambda p_\lambda p_{\lambda\nu} \log \frac{p_\lambda p_{\lambda\nu}}{\sum_\mu p_\mu p_{\mu\nu}} \dots \dots (IV, 11)$$

The average quantity of information received per symbol is therefore

$$H - \bar{h} = - \sum_\nu \left[ p_\nu \log p_\nu - \sum_\lambda p_\lambda p_{\lambda\nu} \log \frac{p_\lambda p_{\lambda\nu}}{\sum_\mu p_\mu p_{\mu\nu}} \right] \dots \dots (IV, 12)$$

The noise properties of the communication channel are entirely determined by the transition probabilities  $p_{\mu\nu}$ . For a noise-free channel, to which (IV, 7) applies, it can be readily verified that

$\bar{h} = 0$ . If the transition probabilities are given, it is possible to choose the relative probabilities  $p_v$  of the code symbols to be transmitted in such a way that the symbols most liable to interference (with the highest  $h_v$ -values) are least used. This means a special choice of coding. With an optimum choice of coding,  $\bar{h}$  is small and the quantity of information transmitted per symbol is maximum. The quantity of information then transmitted is called the *capacity* of the communication channel. Shannon has demonstrated that the capacity of a communication channel can in principle be fully utilized by dividing each message into long series of symbols (letters) and representing each series by a unique set of code symbols (series of pulses). Of all sets of symbols that it would be possible to use, only those are chosen for the code that differ so much that they are least liable to confusion. In practice this would require a sort of dictionary of all kinds of extensive messages, ready for use and translated into sets of code symbols. The scope of the code-books and the delay in looking for the "translation" obviously impose a limit to the possibility of realizing this ideal in practice.

We shall now apply the foregoing to the earlier-described simple model of a communication system, in which coding is effected in the form of voltage pulses whose height determines the symbol. In a noise free communication channel an infinite number of symbols is in principle available, since both at the transmitting and at the receiving end a clear distinction can be made between arbitrarily small differences in amplitude. In this way any message could be expressed in a single pulse by measuring its height to millions of decimal places, the first two decimals representing the first letter, the next two decimals the second letter, etc. In reality, noise is always present, with the result that the original height of a transmitted pulse can be determined at the receiving end only with a given degree of uncertainty, so that only pulses that already possessed a finite difference in height at the transmitting end can be distinguished.

Thermal noise has, at constant temperature and in linear networks, a stationary character, i.e. the probability distribution of the fluctuating noise voltage  $V$  does not depend on time. This probability distribution is Gaussian: the probability that the voltage at a given moment lies between  $V$  and  $V + dV$  is given by

$$W(V) dV = \sqrt{\frac{a}{\pi}} e^{-aV^2} dV. \quad (\text{IV}, 13)$$

The root mean square of the noise voltage, or the

power developed in a resistance of  $1\Omega$ , then amounts to:

$$N = \overline{V^2} = \int_{-\infty}^{+\infty} V^2 e^{-aV^2} dV = \frac{1}{2a}. \quad (\text{IV}, 14)$$

Two transmitted pulses of different height can only be distinguished at the receiving end if the difference in level, expressed as a measure of energy, is at least of the same order as the amount given by (14). A further calculation shows that if the voltage levels differ by an amount  $\sqrt{2N} = \sqrt{1/a}$ , there is probability of about  $\frac{1}{2}$  (0.48 to be exact) that noise will cause pulse distortion to the extent of half the distance to an adjacent level. There is thus a chance of 0.48 that a transmitted pulse will be faultily received (too high or too low). The levels should therefore be placed somewhat further apart. If intervals of  $\sqrt{6N}$  are used, the chance of faulty reception becomes 0.22, which is considered acceptable in practice. For a communication system of the type described here the levels will, therefore, be placed at voltage intervals of  $\sqrt{6N}$ . Only a finite number of pulse heights is then, of course, permissible, viz.  $\sqrt{E/N} + 1 = \sqrt{s} + 1$ ,  $E$  representing the mean signal power, and hence  $s$  the signal-to-noise ratio (in terms of energy).

The above number is arrived at as follows. Let the maximum pulse voltage be  $V_{\max}$ . If the number of equidistant pulse heights is  $n + 1$ , then the various pulse amplitudes are given by  $gV_{\max}/n$ , where  $g$  may have any integral value between 0 and  $n$ . Each pulse represents a power  $\frac{1}{2}g^2V_{\max}^2/n^2$ , and if all pulse heights occur with equal probability, the mean signal power amounts to

$$E = \frac{1}{2} \frac{V_{\max}^2}{n^2} \frac{1^2 + 2^2 + \dots + n^2}{n + 1} \approx \frac{V_{\max}^2}{6}.$$

Hence  $V_{\max} \approx \sqrt{6E}$ . Since the distance between the levels was fixed at  $\sqrt{6N}$ , the number of permissible levels becomes

$$\sqrt{\frac{6E}{6N}} + 1 = \sqrt{\frac{E}{N}} + 1 = \sqrt{s} + 1.$$

In practice not only are the pulse heights restricted but also the pulse frequency, i.e. the number of pulses that can be distinguished per second. This is attributable to the fact that in every communication channel only a limited frequency range is available for the transmission. Let the available bandwidth be  $f_0$ ; in other words, assume that only frequencies between a given frequency  $f_c$  and the frequency  $f_c + f_0$  are transmitted. It can be proved on general grounds that two signals each of duration  $\tau$  seconds can no longer be distinguished if they have the same amplitude at more than  $2f_0\tau$  instants.

A periodic signal of period  $\tau$  can be represented by the following Fourier-series:

$$v(t) = \frac{a_0}{2} + \sum_1^{\infty} a_k \cos 2\pi k \frac{t}{\tau} + \sum_1^{\infty} b_k \sin 2\pi k \frac{t}{\tau}$$

If only the frequencies between 0 and  $f_0 = 1/\tau$  occur, only  $2l + 1 = 2f_0\tau + 1$  terms occur in these expansions, and  $v(t)$  is thus completely determined by the  $2l + 1$  available coefficients ( $a_0, a_1, \dots, a_l, b_1, \dots, b_l$ ). More coefficients do not define the shape of the signal any better. The  $l$  may be neglected if  $2f_0\tau \gg 1$  which is always the case. For a non-periodic signal of duration  $\tau$  the series is replaced by a Fourier-integral, but the conclusion that  $v(t)$  is completely defined by  $2f_0$  data per second remains valid. These conclusions, drawn for the frequency range between 0 and  $f_0$ , are equally applicable to the arbitrary frequency range between  $f_c$  and  $f_c + f_0$ .

Pulse series lasting  $\tau$  seconds are thus completely defined by the amplitude at  $2f_0\tau$  instants. This means that the maximum quantity of information that can be contained within a bandwidth of  $f_0$  c/s in a pulse series of duration  $\tau$  seconds is attained if in that time  $2f_0\tau$  pulses are transmitted.

Since, as we have shown above, a choice can be made of  $\sqrt{s} + 1$  discrete heights for the transmission of each pulse, the total number of distinct (i.e. different) pulse series of length  $\tau$  amounts to  $(\sqrt{s} + 1)^{2f_0\tau}$ . In such a series of length  $\tau$ , therefore, the maximum quantity of information is

$$I = 2 f_0 \tau \log_2 (\sqrt{s} + 1) \text{ bits.} \quad (IV, 15)$$

This largest possible quantity of information is what we earlier termed the channel capacity; this time, however, it is expressed in terms of the bandwidth and the signal-to-noise ratio of the communication channel in question.

The derivation given above is broadly that given by Tuller<sup>14</sup>. Shannon<sup>1</sup>) arrives at the directly comparable result:

$$I = f_0 \tau \log_2 (1 + s) \quad (IV, 16)$$

For strong signals this expression is virtually identical with (IV, 15). The relationship between efficiency of information-transmission and bandwidth of the communication channel, mentioned briefly above in connection with the television picture, will now be clear. The greater the redundancy of the message to be transmitted the less efficiently will the channel capacity be utilized, and a bandwidth greater than follows from (IV, 16) will then be required for transmitting a given quantity of information.

#### Information and energy

If a signal is transmitted over a cable and we wish to receive the maximum signal energy, the

cable must have a reflection-free termination. In that case, as mentioned in the first article of this series (Vol. 16, p. 262), a noise power  $N = kT_a f_0$  in a bandwidth  $f_0$  is produced by the cable at a cable temperature  $T_a$ ;  $k$  represents Boltzmann's constant.

If the signal power that can be extracted from the cable amounts to  $S$ , then the maximum quantity of information that can be transmitted in the time  $\tau$  is, according to (IV, 16):

$$I = f_0 \tau \log_2 \left( 1 + \frac{S}{kT_a f_0} \right) \quad (IV, 17)$$

The quantity of information available per unit time and per unit signal power is then:

$$\frac{I}{S\tau} = \frac{f_0}{S} \log_2 \left( 1 + \frac{S}{kT_a f_0} \right) \quad (IV, 18)$$

Considered as a function of the variable  $S/f_0$ , this function always has a negative differential coefficient. Consequently it reaches its maximum in the limiting case where the signal power is zero. By expanding formula (IV, 18) in terms of  $S/kT_a f_0$ , the maximum available quantity of information per unit time and per unit power is found to be:

$$\left( \frac{I}{S\tau} \right)_{\max} = \frac{1}{kT_a} \log_2 e \quad (IV, 19)$$

Thus, to gain 1 bit of information per second we require at the very least (i.e. for very small  $S/f_0$ ) an energy of  $kT_a / \log_2 e$ , i.e.  $kT_a \ln 2$ . This consequence of Shannon's calculations has been pointed out by Felker<sup>15</sup>). According as the bandwidth is smaller and the signal power greater, the energy required per bit increases.

In this connection, communication theory provides a means of solving a classical paradox relating to the second law of thermodynamics, viz. that concerning "Maxwell's demon". Maxwell had pointed out that a being capable of acting on a molecular scale would be able to cause changes in state that are in direct conflict with the second law. Let us imagine, e.g. a gas-filled vessel whose walls are perfectly insulating. A partition, also a perfect insulator, with a small hole in it divides the vessel into two compartments. The "demon" can close this hole with a valve, which he operates in such a way that only rapid molecules are passed from right to left and slow ones from left to right. In this manner he divides the gas into a hot and a cold part, thus reducing the statistical-thermodynamical entropy of the system without doing any work. It has often been pointed out that this reasoning is unsatisfactory, for several reasons.

<sup>14</sup>) W. G. Tuller, Proc. Inst. Radio. Engrs. 37, 468-478, 1949.

<sup>15</sup>) J. H. Felker, Proc. Instn. Radio Engrs. 40, 728-729, 1953.

Brillouin in particular has drawn attention to the fact that the demon would require information regarding the speed of the molecules for his manipulations. The demon would be unable to "see" anything in a space of constant temperature and would have to use the additional energy of a "lamp". This information would cost at least as much energy as corresponds to the gain in entropy — and, in practice, far more, even with the most favourable measuring technique conceivable.

It is interesting from a historical point of view that as early as 1929 Szilard<sup>16)</sup>, on thermodynamical considerations, came to the conclusion that 1 bit of information (though he expressed it otherwise) in a system of constant temperature would have to cost an energy of at least  $kT \ln 2$  in order to avoid conflict with the second law. Szilard's imaginary experiment can be briefly described as follows. A single molecule is contained in a vessel, kept at a constant temperature by a heat reservoir. In the vessel is a piston with an opening that can be closed with a sliding valve. The piston is initially in the middle of the vessel. At the beginning of the experiment the observer closes the opening and looks to see whether the molecule is at the right or at the left. Then he allows the piston to be gradually moved in the opposite direction. The molecule hits the piston several times, thus expending work upon it. This is the energy which corresponds to the isothermal expansion of a perfect gas from the volume  $\frac{1}{2}V$  to the volume  $V$ , during which an amount of heat  $kT \ln 2$  per molecule is abstracted from the reservoir (This is deducible from I, equation 15). After some time the piston reaches its extreme position. The valve is now opened and the piston is returned to its initial position, after which the whole process can be repeated.

At first sight the foregoing conflicts with the second law, since energy is being continuously extracted from a constant temperature source apparently without doing any work. In order to avoid such a discrepancy, however, Szilard assumes that the measurement itself, which involves determining the place of the molecule — i.e. supplying one bit of information (right or left, i.e. two equally probable possibilities). — requires an energy of at least  $kT \ln 2$ . The application of ideas such as entropy, temperature, isothermal expansion etc. to this gas, which consists of one molecule, as well as the idea of action on a molecular scale in this imaginary experiment deserves to be subjected to criticism. Szilard himself partly answers this

criticism by extending the argument to large numbers of molecules; others such as Raymond, have since described analogous imaginary experiments with gases consisting of more than one molecule. However, the merit of Szilard's achievement remains: he derived an interesting result which satisfactorily agrees with present ideas in information theory.

In this article we have seen that there is a connection between the "entropy" of information theory, and the statistical-thermodynamical entropy. We found this connection when determining the amount of energy needed for the transfer of information through a channel with *noise*. Since the intensity of the noise depends on the temperature, the energy must also depend on temperature, and it is therefore hardly surprising that the thermodynamical notion of entropy should come into these considerations. Discussion of Szilard's experiment, however, shows how little the term "entropy" is justified in information theory. The use of this term has sometimes been responsible for the following explanation of the result obtained by Szilard: 1 bit of "information entropy" equals a "thermodynamical entropy" of  $k \ln 2$ . However, the introduction of Boltzmann's constant  $k$  in pure information theory is really quite irrational (as opposed to its natural appearance in *thermodynamical* considerations of communication channels, etc.). For example, what could be the significance of Boltzmann's constant when dealing with the "entropy" of a language, where there is no question of either noise temperature or energy? What indeed could be the thermodynamical "system" of which the language forms a part? In the absence of such a system, Boltzmann's constant clearly has no place.

So long as we confine ourselves to the statistical aspect of the notion of entropy there is no danger of confusing thermodynamics and information theory. The mathematical analogy, however, does not imply that the experimental laws which are valid for thermodynamical entropy (second law), should have a corresponding physical significance in information theory.

**Summary.** In the relatively new branch of science known as information theory, which is mainly concerned with communication and cybernetics, the notion of information is given a quantitative meaning on the basis of statistical considerations. A quantity termed "information content per symbol" is thus derived, which is formally closely analogous to statistical-thermodynamical entropy (see the previous articles I, II and III of this series). The concept can be easily applied to any given series of events each of which may occur with a certain degree of probability, as for example the occurrence of letters in a language. This has led some to speak of the "entropy" of a language, or the "entropy" of a television picture.

<sup>16)</sup> L. Szilard, Z. Phys. 53, 840-846, 1929.

An important application of information theory lies in the investigation of the efficiency of the transmission of information, e.g. by electric signals. By methods which show a certain formal agreement with those applied in statistical thermodynamics, it is possible to derive highly general results concerning the way in which the noise and bandwidth of the communication channel affect the capacity of the channel. Finally, the minimum quantity of energy needed for the transmission of a given quantity of information per unit time is

discussed. Owing to the effect of thermal noise, it is found to depend on the temperature of the communication channel. This dependence on temperature establishes a connection with thermodynamical entropy, but the connection is based on trivial grounds, and constitutes no justification whatsoever for the common use of the term "entropy" in the theory of information. This abuse may lead us wrongly to expect analogies between thermodynamics and the information theory which go beyond their common statistical background.

## ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk \* can be obtained free of charge upon application to the Philips Research laboratory, Eindhoven, Netherlands.

**2330:** A. M. Kruihof: Quelques idées sur les transformations du verre (Verres et Réfractaires 9, 311-319, 1955).

Measurements made with different glasses show that any glass can have various densities at room temperature. The difference in density is the result of different thermal histories of the glass samples. On the basis of a theory of the freezing-in of structure states, the existence of different densities at room temperature can be explained with the use of the contraction curves. Following the definition of Tool, these structure states can be characterized by a "fictive temperature". Ideas on changes in structure states can explain the mutual differences between expansion curves as well as differences between these curves and the contraction curves. The conceptions of "transformation point" and "transformation range" are discussed on the basis of this theory. Other physical properties e.g. viscosity and electrical conductivity (direct current) show similar phenomena. The curve representing the log of the specific resistance as a function of the inverse of the absolute temperature consists of three straight lines. Both points of intersection may correspond to transformations in the glass. Thermal differential analysis was used as a check on this hypothesis. That the lower point does correspond to a transformation is not, however, completely established.

**2331:** H. G. van Bueren: Theory of the formation of lattice defects during plastic strain (Acta Met. 3, 519-524, 1955).

A simple theory is presented by which a relation between the plastic strain and the concentration of vacancies, interstitials and dislocations is obtained. Dislocations are formed by sources under the action of an applied stress, the other defects by the move-

ment of jogs in the dislocations. The action of a dislocation source under a varying stress is studied, for the case of static as well as of dynamic generation. After the first few percent of strain, both methods of generation yield the same resulting defect concentrations. By suitable elimination, the influence of work hardening could be left out of the theory, and the defect concentrations could be expressed as functions of the amount of plastic strain. The theoretical deductions should be valid between plastic strains of 0.05 up to 1. They are compared with the observed critical shear stress, the elementary structure and the resistivity-strain relation in slightly deformed copper.

**2332:** G. Brouwer: Electrical analog of the eddy-current-limited domain-boundary motion in ferromagnetics (J. appl. Phys. 26, 1297-1301 1955).

The observed losses in a ferromagnetic core are always greater than those calculated on the basis of a homogeneous permeability. In a magnetization process by domain-boundary displacements the permeability certainly is not homogeneous, but attains extreme values in the domain walls. As a result the loss factor due to eddy-currents is increased. This is the well-known eddy-current anomaly. The effect increases with the domain size. An electric circuit analogue was used to determine the eddy current distribution and domain-boundary motion in a number of idealized cases.

**2333\*:** A. van Weel: Un nouveau système de mesure des angles de phase (Aéroélectronique 1er Congrès International, Dunot, Paris 1955, pp. 565-572).

If a four-pole is inserted in the feedback loop of an oscillator, the frequency of the latter is affected

by the phase change suffered by the waveform on passing through the four-pole (See also Philips tech. Rev. 15, 307-316, 1953/54.) This principle offers considerable advantages for phase angle measurement. By transforming phase variations into frequency variations, phase meters with the following properties can be designed: range 0-360° or more, precision 1°; small variations of phase measurable to an accuracy better than 0.01°; direct indication on a calibrated linear scale. For certain applications the unknown phase angle originates in a four-pole which cannot be inserted in the oscillator loop. To measure the phase relation between two A.C. voltages a new principle has been devised which permits the unknown phase angle to be introduced into a separate oscillator of special configuration. To the above-mentioned properties may then be added: insensitivity to not-too-large frequency variations in the given voltages. Possible applications include altimeters and rangefinders.

- 2334: H. B. G. Casimir: On the theory of superconductivity (Reprinted from: Niels Bohr and the development of physics, Pergamon Press, London 1955).

Survey of certain aspects of superconductivity and of proposed theories of this phenomenon.

- 2335: J. de Jonge and B. H. Bibo: The preparation of ortho and para hydroxybenzyl alkyl ethers (Rec. trav. Chim. Pays Bas 74, 1448-1452, 1955).

Alkyl ethers of ortho and para hydroxybenzyl-alcohol can be obtained with a good yield by heating the hydroxybenzyl alcohol with the appropriate alkyl alcohol at 150 °C.

- 2336: G. Thirup: Wide-band three-phase RC-generators for complex measurements of two-poles and four-poles (J. Brit. Instn. Rad. Engrs. 15, 597-605, 1955).

From a 3-phase R-C oscillator two voltages are derived, the complex ratio of which can be varied. The complex ratio is independent of the frequency. The two voltages are used in a composition circuit for measuring the parameters of two-poles and four-poles. Two equipments are described covering the frequency ranges 20 c/s-22 kc/s and 22 kc/s-10 Mc/s. The possible error is  $\pm 0.5$  db and  $\pm 2^\circ$  in the frequency range 100 c/s — 3 Mc/s. Outside this range the phase error may increase about 3 times while the amplitude error remains nearly the same.

- 2337: P. Jongenburger: The magneto-resistance of metals deformed at low temperatures (Suppl. Bull. Inst. Int. du Froid, Annexe 1955).

Carefully annealed pure copper wires were plastically deformed at 20 °K. Their magneto-resistance was measured at the same temperature before and after the deformation. When the results were plotted in a Kohler diagram, a large influence of the deformation was observed. This effect was shown to be due to dislocations only; it was independent of the simultaneous presence of point-defects.

- 2338: F. A. Kröger: The physical chemistry of crystal phosphors (Proc. I.R.E. 43, 1941-1944, 1955, No. 12).

After an historical introduction, a survey is given of the present views regarding the constitution and the preparation of inorganic crystals phosphors. Particular attention is paid to the incorporation of atoms with a valency deviating from that of the atoms of the base material, and to the stabilization of atoms in a particular valency.

- 2339: E. W. Gorter: Some properties of ferrites in connection with their chemistry (Proc. I.R.E. 43, 1945-1973, 1955, No. 12).

After an elementary introduction on the origin of the magnetism of oxides, it is shown how the molecular-field hypothesis can amount for the magnetic properties of ferromagnetics and antiferromagnetics, and for those of non-compensated antiferromagnetics, with which latter materials we are concerned here. A brief description of the spinel lattice is given, and also an account of the crystal chemistry of the spinels, which is necessary to understand the experimental saturation magnetizations discussed. A short survey of methods of preparation is given. The second part discusses the anisotropies, and some of the magnetization processes which influence permeability, and the factors which influence high-frequency permeability and losses. Among these are the ferromagnetic resonance phenomenon and the dimensional resonance and relaxation phenomena. The way in which these factors are influenced by chemical composition and preparation technique is indicated. Finally, a short history of the development of ferrites is given.

- 2340: C. O. Jonkers: Die Philips Gaskältemaschine (Allgemeine Wärmetechnik 6, 203-206, 1955).

Abbreviated version (in German) of article published in Philips tech. Rev. 16, 69-78, and 105-115, 1954/55.

- 2341:** J. S. van Wieringen: Paramagnetic resonance of divalent manganese incorporated in various lattices (Disc. Faraday Soc. No. 19, 1955).

Paramagnetic resonance was observed in powder samples containing Mn diluted in various diamagnetic compounds. The measurements were made at 3.2 and 1.25 cm, at room temperature. Three effects were found: narrowing of hyperfine splitting by covalent bonding and by exchange, and a  $g$ -factor somewhat larger than the free-electron value in ZnSe-MnSe and CdTe-MnTe mixtures. The first effect suggests 10-20%, 30%, 35% and 40% covalent bonding in Mn-oxygen compounds, MnS, MnSe and MnTe respectively.

- 2342:** A. A. Kruithof and J. L. Ouweltjes: Colour rendering by de luxe fluorescent lamps (Proc. Int. Comm. Illn. Zürich, 1955).

Description of a method of improving the colour rendering of fluorescent lamps by the addition of a special phosphor. The principles governing the colour rendering are discussed and the application of these principles demonstrated with regard to some newly developed "de luxe" lamps.

- 2343:** J. J. Balder and G. J. Fortuin: The influence of time of observation on the visibility of stationary objects (Proc. Int. Comm. Illn. Zürich, 1955).

Interim report of new measurements, with new apparatus, of the threshold values of visual acuity. Apart from the variables concerned in an earlier investigation by Fortuin (see these abstracts R 170 and R 174), the effect of object observation time was also examined.

- 2344:** H. A. Klasens, P. Zalm and G. Diemer: Characteristics of electroluminescent cells (Proc. Int. Comm. Illn. Zürich, 1955).

A consideration of the chemical nature of an electroluminescent phosphor is followed by a discussion of the electrical and optical characteristics of an electro-luminescent cell, making use of a simple equivalent circuit. On the basis of these considerations, and direct observations, the properties of the individual phosphor crystals are derived. A relation is found expressing the brightness  $B$  in terms of the frequency  $\omega$  and the applied voltage  $V$ . A theory is proposed for the mechanism of electroluminescence which explains this relation. Finally, some comments are made on the practical application of electroluminescent lamps.

- 2345:** H. Zijl: Computed coefficients of utilization (Proc. Int. Comm. Illn. Zürich, 1955).

A method is described by which utilization tables can be computed which take account of the nature of the illumination of the working surface and the light distribution over all the surfaces enclosing the illuminated space. It is possible to present the data in the form of simple graphs and tables.

- 2346:** J. B. de Boer and J. F. T. van Heemskerck Veeckens: Observations on discomfort glare in street-lighting; influence of the colour of the light (Proc. Int. Comm. Illn. Zürich, 1955).

Experimental investigation into the effect of the colour of the light on glare in street lighting. Of the various lamps used, the observers showed a definite preference for the "warmer"-coloured light sources. The acceptable luminance (to ensure a certain degree of comfort) was found to be approximately proportional to the square root of the road luminance. Differences in the results from those of other investigators can partly be attributed to differences in the measuring techniques.

- 2347:** H. Bremmer: Diffraction problems of microwave optics (I.R.E. Trans. 4, 1955).

Survey of current methods applied in microwave diffraction theory.

- 2348:** A. A. Kruithof: Chromatic adaption with near white backgrounds (Die Farbe 4, 147-158, 1955).

Repetition and extension of work carried out earlier in collaboration with P. J. Bouma (Philips tech. Rev. 9, 257-266, 1947/48).

- 2349:** J. L. Meijering, G. W. Rathenau, M. G. van der Steeg and P. B. Braun, A miscibility gap in the face-centred cubic phase of the copper-nickel-chromium system (J. Inst. Metals 84, 118-120, 1955/56).

The copper-nickel-chromium system has been studied by metallographic and X-ray-diffraction methods, particular attention being paid to the isothermal section at 930 °C. In addition to the body-centred cubic phase rich in chromium, two face-centred cubic phases have been found to exist, although the binary copper-nickel alloys show complete miscibility.

- 2350:** C. Meyer: Gasentladungslampen und ihre physikalischen Probleme (Physikalische Blätter 12, 19-28 and 62-73, 1956). (Gas discharge lamps and their physical problems; in German)

Following a short historical review, questions relating to the efficiency of luminous sources,

fluorescence, colour rendering and the conditions of operation (striking and stabilization) are discussed. The lamps discussed include sodium vapour lamps, high pressure and super-high pressure mercury lamps, high pressure xenon lamps and electronic flash tubes. The article ends with a table showing the spectral distribution of the light from a number of gas-discharge lamps as compared with daylight and incandescent lamplight.

- 2351: O. W. Memelink: The distribution of impurity in a semi-infinite solidified melt (Proc. Phys. Soc. B 69, 119-120, 1956).

The transient distribution of impurity segregating upon solidification in a semi-infinite melt has been calculated. The expression obtained, which is of closed form, contains the following parameters: the constant velocity of the moving solid-liquid interface, the ratio of impurity solubility in the solid to that in the liquid state and the diffusion constant of the impurity in the liquid state.

- 2352: J. E. Rombouts and J. Links: The chemical nature of the antibacterial substance present in *Aucuba japonica* Thunbg (Experientia 12, 78-80, 1956).

An inactive precursor of the antibacterial substance present in the aucuba plant has been isolated and identified from the seeds. The precursor was identified with aucubin, a glycoside of a furan derivative. From the leaves an enzyme could be extracted which produced from the aucubin (presumably by hydrolysis) the antibacterial substance aucubigenin. The activity spectrum of this substance is described. It is asserted that the antibacterial substance present in crude juices from leaves is identical with the hydrolysis product of aucubin.

- 2353: H. J. G. Meyer: Interaction of excitons with lattice vibrations in polar crystals I (Physica 22, 109-120, 1956).

A theory describing slow excitons which interact with longitudinal optical lattice vibrations is developed within the framework of the effective mass approximation. In many respects the theory is very similar to — and makes explicit use of — the ordinary theory of polarons. Interesting differences with polaron theory are introduced by the fact that an exciton has an internal degree of freedom and is as a whole electrically neutral. The effect of lattice vibrations on optical exciton transitions is investigated with the aid of the above theory.

- 2354: G. D. Rieck: Fragmentation in tungsten crystals (Acta Met. 4, 47-51, 1956).

Back-reflexion-Laue diagrams were made of W crystals in recrystallized lamp filament wires, using a specially collimated X-ray micro-beam. Differences in orientation of 1-2° could be observed. Sometimes two crystals with orientation difference of 1°-2° were irradiated at the same time. The crystals often showed Laue diagrams with spots split in the direction of the wire axis, with angular differences of 2'-30'. After cold bending, fragmentation was found, which was more clearly seen after re-straightening, for then the deformation asterism diminished. The fragmentation occurred mainly in two directions. One corresponds with the normal breaking-up of the lattice in the direction of bending. The other was approximately perpendicular to the latter and gave small lines parallel to the wire axis, arising from particles with orientation differences of 1'-10' and dimensions of about 10 μ. Bearing in mind the split spots occurring in unbent wires and the fact that during drawing contaminations in the wire are stretched out along its length, we may assume a predisposition of the crystals to break along planes parallel to the axis.

- 2355: W. K. Westmijze: The fundamentals of magnetic recording (T. Ned. Radiogenootschap 21, 1-15, 1956).

A short description is given of the magnetic recording method in general, and some details are treated more elaborately. For the understanding of the h.f. biasing method use is made of a simplified hysteresis curve. In this way an explanation can be given of some of the peculiarities met in recording and erasing. The recorded signal is attenuated by the demagnetizing field, and during reproduction not all the flux in the tape is reproduced. This is discussed for short, long and intermediate wavelengths. Finally, the factors are surveyed which influence distortion, frequency response, noise and print effect.

- 2356: G. P. Bakos: Mechanical aspects of magnetic-recorder design (T. Ned. Radiogenootschap 21, 17-37, 1956).

The quality of magnetic recording is determined to a great extent by the electro-mechanical design of the tape-drive mechanism. The task of this mechanism is to make the tape run with a constant speed and in intimate contact with the magnetic heads. It depends on the constructional details how far this task is fulfilled. This article deals with the various factors which have to be taken into account when designing tape drive mechanisms.

# Philips Technical Review

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

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

## PROBLEMS IN THE CONSTRUCTION OF SMALL RADIO VALVES

by B. A. CANT \*).

621.385.1-181.4:621.396.694

*In the last 20-25 years radio valves have been made smaller and smaller. The sub-miniature types, evolved for use in portable apparatus, supplied by battery, are extremely small. Although the advent of the transistor in the last few years has meant the replacement of these valves in some fields of application, there remain many instances in which the transistor cannot yet be substituted for the radio valve. In this article the author will deal with some of the problems encountered in the construction of valves of very small dimensions.*

### Introduction

The earliest radio valves were used in rather bulky receivers, so that their size was immaterial. With the increasingly higher demands imposed upon valves, the dimensions of newly developed types became smaller and smaller. This trend has often been described in this Review <sup>1)</sup>. It is particularly apparent in the domain of battery-fed valves, used predominantly in portable sets, for which small dimensions and light weight are prime considerations. Whereas the envelope of a battery-fed valve such as the KF 3 had a diameter of 40 mm, this dimension is only 19 mm for valves like the DF 96 and DF 97. The two latter belong to the series of so-called *miniature valves*. More recently valve dimensions were even further reduced in a series of *sub-miniature valves* (fig. 1), which found application in hearing aids <sup>2)</sup>, for example. The rapid development of the transistor has diminished the importance of valve-equipped hearing aids, but nevertheless large numbers of these units are still in use. Sub-miniature valves are also used in other

equipment, such as radio sondes <sup>3)</sup> and small portable transceivers (walkie-talkies). For many functions in such equipments, radio valves can as yet not be replaced by transistors, mainly owing to the fact that valves can be used at higher frequencies than currently available transistors.

In this article we shall deal with some of the problems in the construction of miniature and sub-miniature valves that have claimed particular attention. Since we are nearly always concerned with valves for battery supply, we shall also pay attention to the various measures taken to reduce their current consumption to a minimum.

### Shape and dimensions of the envelope

The construction of the first receiver valves was simply modelled on the incandescent lamp. The electrodes were housed in a fairly large glass bulb, and a number of lead-in wires, held together in a pinch, formed the connection to the electrodes. Fig. 2a shows how the bulb of such a valve with pinch construction was sealed off. The valves were then finished off by fitting them with a bakelite base with pressed-in contacts. The drawbacks of this construction have been pointed out in the articles quoted in <sup>1)</sup>. In a later type, of "all-glass" con-

\* ) The Mullard Radio Valve Co., Ltd., Mitcham, England.

<sup>1)</sup> P. G. Cath, A new principle of construction for radio valves, Philips tech. Rev. 4, 162-166, 1939. Th. P. Tromp, Technical problems in the construction of radio valves, Philips tech. Rev. 6, 317-323, 1941. G. Alma and F. Prakke, A new series of small radio valves, Philips tech. Rev. 8, 289-295, 1946. J. L. H. Jonker, Electron tubes, Philips tech. Rev. 14, 117-128, 1952/53.

<sup>2)</sup> Cf. P. Blom, An electronic hearing aid, Philips tech. Rev. 15, 37-48, 1953/54.

<sup>3)</sup> See A. Hauer and M. van Tol, A radio sonde for meteorological observations, Philips tech. Rev. 16, 148-156, 1954/55.

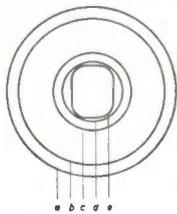
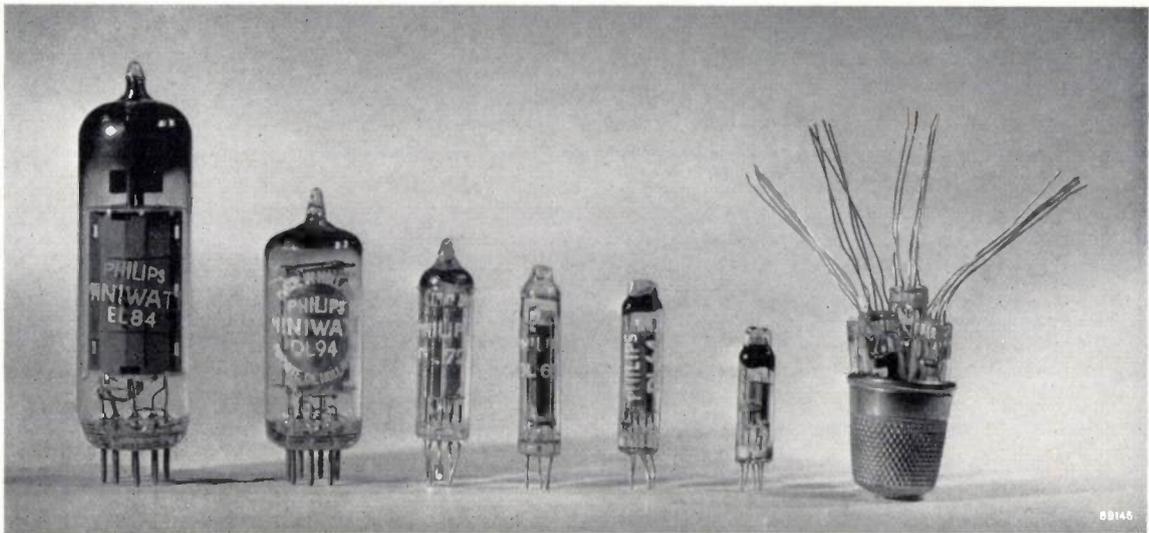


Fig. 1. Some sub-miniature valves compared with two larger valves. The photo shows from the left to the right: an indirectly-heated power amplifier pentode (EL 84), a miniature valve (DL 94), the round sub-miniature valves DL 72 (diameter 10 mm) and DL 65 (diameter 7.9 mm), the flat sub-miniature valves DL 64 and DF 64, and finally four DL 64 valves fitted in a thimble.

The inset sketch shows the bulb cross-sections of these valves, drawn full size. *a*) EL 84 (diameter 22 mm), *b*) DL 94 (19 mm), *c*) DL 72 (10 mm), *d*) DL 65 (7.9 mm), *e*) DL 64 and DF 64 (7.25 × 5.6 mm).

struction, both cap and pinch were eliminated, and this considerably improved the performance of the valve in several respects. The dimensions of the valves manufactured on this principle became smaller and smaller. The envelope diameter of the miniature valves, as mentioned earlier, was 19 mm. A subsequently developed "all-glass" series, the *round* sub-miniature valves, had a diameter of only 10 mm, and even 7.9 mm for some types.

With miniature and round sub-miniature valves, like the older "all-glass" valves, the electrode assembly is mounted on a number of pins sealed into a glass plate; the latter, on being sealed to the envelope, forms the base of the valve. Figs. 2*b* and *c* show the method of sealing-in. Because of the small dimensions, the electrodes are close to the sealing joint and run the risk of becoming so hot during the sealing process that they are liable to mechanical or chemical damage. This risk is even greater for another series of valves, developed by some manufacturers round about the same time as the round sub-miniature valves. Here the envelope is of a rectangular cross-section (some types having dimensions of 7.25 × 5.6 mm) which, for rectangular or oval-shaped electrodes, makes possible a more efficient use of the space available in the envelope. Sealing of the envelope is effected by pressing its outer edges together over the connecting wires, the latter being arranged in a single plane (fig. 2*d*). Here, in fact, we once again have the pinch type of construction with its attendant drawbacks.

Indeed, owing to the small dimensions and to the different manufacturing procedure, some difficulties are encountered that were either absent or present to a much slighter degree in the older valves with pinch construction. Whereas with the old valves the pinch was made before the electrodes were connected to the lead-in wires, with subminiature valves the operation is not performed until the whole electrode assembly has been completed and mounted on the connecting wires. Since during sealing-off the glass is heated up to 800-900 °C, there is a considerable risk of the electrodes being subjected to inadmissibly high temperatures. The danger is greatest for the cathode, whose emissive power may be gravely impaired by excessive heating during sealing-off. It will be clear that, with a construction in accordance with figs. 2*b* and *c*, the electrodes are far less liable to excessive heating.

Another disadvantage of the construction of fig. 2*d* is that the lead-in wires have to be fixed at definite intervals and in the same plane whilst being sealed into the envelope. This difficulty does not occur with a construction where the lead-in pins are fixed in a glass base plate before the electrodes are mounted and before sealing-in, as in the miniature and round sub-miniature valves.

In a new series of sub-miniature tubes it has been found possible to combine the advantages of an envelope of rectangular section with those of electrode mounting on a flat base plate. The latter is oval in shape and contains, like that of the minia-

ture and round sub-miniature valves, the lead-in wires on which the electrodes are mounted before the valve is sealed-off.

Some additional provisions are made to prevent overheating of the electrodes during sealing-off. The envelope, for instance, is placed with the base plate on top, and a stream of an inert gas is blown through the envelope during sealing. The rim of the base plate is kept as thin as possible, so that a short heating time will be sufficient. The gas jets, moreover, are of such a shape and size that heating is strictly localized, whilst penetration of the flames

containing a small proportion of other elements was used, both for directly and indirectly heated cathodes, as it forms a better basis for the emissive oxide layer that has to be deposited on the cathode. The endeavour to restrict the filament current of battery-fed valves to the barest minimum, however, necessitated a return to tungsten as filament material. The fairly low tensile strength of nickel wire did not permit the manufacture of wire thinner than 22-24  $\mu$ . Tungsten wire, having a greater tensile strength, has been manufactured in filaments of only 6.3  $\mu$  for use in some sub-miniature valves,

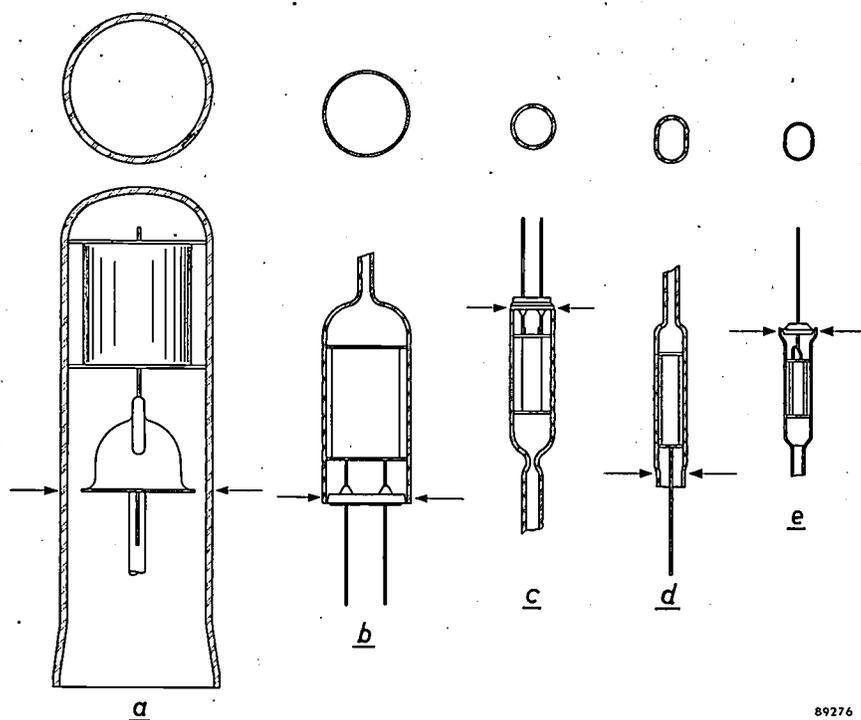


Fig. 2. Methods of sealing-in of various types of electronic valves. The arrows show how the gas flames are applied.

- a) Earlier valve types with pinch construction.
- b) Miniature valves; glass base.
- c) Round sub-miniature valves; glass base.
- d) Flat sub-miniature valves with pinch construction.
- e) Flat sub-miniature valves with flat glass base.

into the envelope is prevented by giving the latter a cup-shaped extremity which accurately fits round the base plate.

Since with this sealing method the cathode is far less liable to overheating, it was found possible to manufacture "rectangular" or "flat" sub-miniature valves with a thinner filament than could be achieved with the pinch construction. Filament-current consumption can thus be substantially reduced.

#### The filament

In the earliest type of radio valves tungsten was used for the (directly heated) filament. Later nickel

making it possible to reduce their filament current to about 10 mA. Another article in this issue <sup>4)</sup> gives further particulars regarding the manufacture of very thin wires of this kind.

The choice of filament diameter is closely connected with the desired temperature. Where battery-fed valves are concerned, several factors must be considered in determining the optimum temperature. At too high a temperature the valve will operate better to start with, but at the sacrifice of

<sup>4)</sup> L. Schultink and P. G. van Zanten, Thin tungsten wire for small radio valves, Philips tech. Rev. 18, 222-228, 1956/57 (No. 8).

its service life. Too low a temperature has the great drawback that the valve properties are unduly affected by the battery voltage. Particularly in battery-fed valves, the filament voltage may drop appreciably below the rated value, and it is naturally desirable that the valves should continue to operate properly even in these conditions.

If in a valve the filament current is not large with respect to the emission current of the cathode, the filament temperature is affected by the latter in two ways. Electron emission itself has a cooling effect; furthermore, the cathode emission current flows through only a part of the filament, so that this part of the total filament length develops additional heat. This non-uniform heating of the filament is obviously undesirable and imposes, for a given emission current, a lower limit on the filament current and hence on the filament diameter. It may be assumed that the filament current should be at least eight times the emission current, which means that for a filament current of 10 mA an emission current of at the most 1.25 mA is permissible. It is for this reason that in some sub-miniature valves a filament somewhat thicker than the above-mentioned minimum of  $6.3 \mu$  is employed, in order to attain a filament current of 15 mA.

### The grids

The grids of sub-miniature valves also present several constructional problems partly due to their small dimensions, but also arising from the fact that the valves have to operate at particularly low voltages<sup>5)</sup>.

are allowed to reach the screen grid, an undesirable drop in the anode current will result. To prevent this, the potential in the plane of the suppressor grid must be given a value such that the reflected electrons cannot pass the suppressor grid. Since, however, the average speed of the reflected electrons is only slightly lower than that of the electrons moving towards the anode, it is difficult to choose the suppressor potential such that only the forward and not the reflected electrons can pass. In addition many of the forward electrons have suffered a change in direction at the screen grid and some will therefore return to the screen grid before reaching the suppressor. In the quest for an acceptable compromise one of the first objects was to ensure the smallest possible deflection of the electrons when passing the screen grid. A condition for this is that the electrostatic field should remain homogeneous up to very near this grid. An improvement can be brought about by reducing the pitch of this grid<sup>7)</sup> and by employing a low screen-grid voltage. The effect of these two measures on the potential distribution is apparent from a number of cross-sections through equipotential planes shown in *fig. 3*.

Now small pitch and low voltage of the screen grid both have their disadvantages. With a small pitch the current to the screen grid may become too great and there is loss in anode slope. For this reason it was necessary to make the screen grid of sub-miniature valves of extremely thin wire. A very low screen-grid voltage can be applied only when the electrode configuration is such that, in spite of the negative bias of the control grid, the average

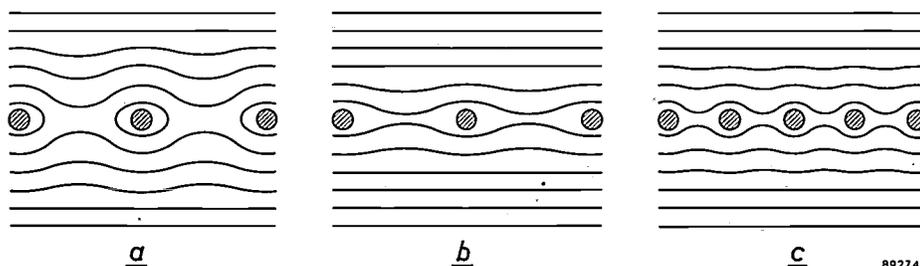


Fig. 3. Cross-sections through some equipotential planes in the vicinity of the screen grid of a pentode. In *b* the screen-grid voltage is lower than in *a*, and in *c* the pitch of the screen grid is lower than in *a*.

At a low anode voltage (c.g. 15 V) the shape of the valve characteristics is greatly affected by electrons reflected from the anode<sup>6)</sup>. If these electrons

potential in the plane of the latter grid remains sufficiently high, for if the potential throughout the whole plane becomes negative (*fig. 4a*), the anode current is suppressed. The "penetrating effect" of the screen grid voltage through the control

<sup>5)</sup> The lower the voltage of a battery, the higher its energy-volume efficiency, i.e. the more watt.seconds per unit volume.

<sup>6)</sup> See J. L. H. Jonker, Reflections in electron tubes, Philips Res. Rep. 2, 331-339, 1947.

<sup>7)</sup> See J. L. H. Jonker, Electron trajectories in multigridded valves, Philips tech. Rev. 5, 131-140, 1940.

grid, therefore, should be made large, which implies that the amplification factor of the control grid with respect to the screen grid should be made small. This could be achieved by increasing the pitch of the control grid (fig. 4*b*), but this would decrease the slope, which is obviously undesirable. However, it was found possible to retain a sufficiently high slope, and at the same time to maintain

A wire-shaped cathode is not very rigid and will readily vibrate. As a rule this will cause a variation in the grid-cathode spacing and consequently a variation in some of the quantities having an important effect on the working of the valve, such as anode current, slope and input capacitance. In a receiver the loudspeaker may convert these variations into air vibrations, either audible or not, which

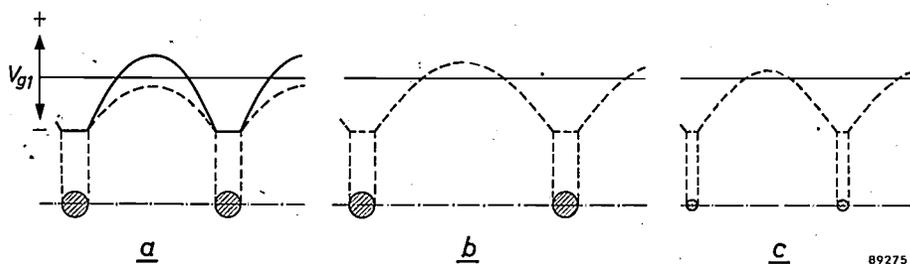


Fig. 4. *a*) The potential in the plane of the control grid. The dotted curve applies to a lower screen-grid voltage than that of the full curve. In the former case the potential in the whole plane is negative.  
*b*) With a larger pitch of the control grid the potential in parts of the control-grid plane may be positive in spite of a low screen-grid voltage.  
*c*) A similar effect is obtainable with thinner grid wires.

a satisfactorily high potential in the plane of the control grid, by using again very thin wire for the winding of the latter (fig. 4*c*).

For a pentode operating at a very low anode voltage the virtually horizontal part of the  $I_a$ - $V_a$  characteristic should preferably extend back as close as possible to the  $I_a$  axis. (A valve used as a voltage amplifier then has a high internal resistance, even for a low  $V_a$ , whilst for a valve used in an output or power stage, where it must supply a certain power, a characteristic of this type makes for high efficiency.) In achieving this the aim should be to minimize the influence of the anode voltage upon the potential in the plane of the suppressor grid. Giving the suppressor grid a small pitch may be expected to help, but this involves the risk of an inadmissible drop of the average potential in the plane of this grid. This in its turn can be prevented by giving the suppressor grid a moderate positive bias. In some sub-miniature types the suppressor grid is joined to the positive end of the filament for that reason.

### Microphony

Since sub-miniature valves are used almost exclusively in small, portable equipments, microphony caused by these valves may be very troublesome. In their design, therefore, special attention should be paid to this problem. As in all other valves with directly heated filaments, the filament itself is the main cause of microphony.

in their turn may react on the valve. In serious cases vibrations thus set up will be maintained as a form of positive acoustic feedback. In less serious cases any knocks against the set can be heard as a damped vibration of the filament.

The intensity of microphony depends on the amplitude of the vibrating cathode. In a valve of symmetrical construction the effect is somewhat compensated by the fact that an increase in distance at one side is accompanied by a decrease at the other side. An exactly symmetrical arrangement of the electrodes, therefore, is essential to combat microphony.

The resonance frequency of a stretched wire is given by

$$f = \frac{1}{2L} \sqrt{\frac{T}{M}},$$

where  $L$  is the length of the wire,  $T$  the tension, and  $M$  the mass of the wire per unit length.

A high resonant frequency (e.g. above 5000 c/s) is desirable for the following reasons:

- The initial amplitude of the damped vibration of the wire resulting from a given impact is inversely proportional to the resonance frequency.
- The feedback of sound vibrations from loudspeaker to cathode is weaker at higher frequencies.
- Very high notes are less audible than those of medium pitch.
- The frequency range of an amplifier or of the audio stage of a radio receiver is limited. If the resonance frequency of the valve filaments lies

above this range, there is but little risk of acoustic feedback.

With a view to combating microphony, therefore,  $T/M$  should be made as large as possible. However, the risk of breaking the wire makes it necessary to limit the tension  $T$  to a value considerably below that at which the hot filament would break (e.g. to about one third of the disruptive force).

Because both  $M$  and the permissible value of  $T$  are proportional to the cross-sectional area of the wire, the maximum attainable resonant frequency depends entirely upon the length. In this respect a short filament, and hence a short electrode system, is desirable.

As a tungsten filament can stand a far higher tension than one of nickel, substantially higher resonant frequencies can be obtained with it. Since, however, the internal damping for mechanical vibrations is greater for nickel than for tungsten, the reverberation time is shorter for a nickel wire.

As an example of what can be achieved to avoid microphony it may be mentioned that the resonance frequency of the filament in the sub-miniature valve DF 64 is as high as 8000 c/s.

Finally it should be pointed out that high demands are imposed, not only upon the filaments themselves, but also upon the springs by which they are stretched, in order to keep the wires under an appropriately reproducible tension.

**Summary.** The general endeavour to make radio valves of ever smaller dimensions presents a number of problems, some of which are dealt with in this article.

For the smallest types of so-called sub-miniature valves the old pinch-construction, long since abandoned in the manufacture of normal radio valves, was first employed. More recently sub-miniature valves have been successfully made using a pinch-less construction. For the filaments and grids of these valves extremely thin wires are used. The filament is made of tungsten instead of nickel, which results in a reduction in filament current. By making the electrode system short and by using high-quality springs for stretching the filament, the resonant frequency of the filament can be raised, which gives a decrease in valve microphony.

## THIN TUNGSTEN WIRE FOR SMALL RADIO VALVES

by L. SCHULTINK and P. G. van ZANTEN.

669.27.426:  
621.385.032.213.13:  
621.375.2:621.396.62

*For the manufacture of miniature and sub-miniature valves extremely thin wires of tungsten and other metals are required, as discussed in the preceding article. The article below gives some particulars regarding the manufacture of such thin wire and deals with a number of requirements it has to satisfy.*

The general aim to reduce the dimensions and the power consumption of radio valves has led to the employment of extremely thin metal wire in the construction of filaments, grids and certain other components. Filaments for sub-miniature and miniature valves, of the directly-heated as well as the indirectly-heated type, are made of *tungsten wire* with a thickness of 50 down to 6  $\mu$  — wire of the same material as used for the filaments of incandescent lamps. Certain grids, particularly the "frame grids" used in some valves for professional application<sup>1)</sup>, are also wound of tungsten wire. Furthermore, tungsten wire is used for some of the smaller components of radio valves, such as hooks and tiny springs for stretching the filaments.

An impression of the extreme thickness of such wire is given in *fig. 1* in which a 15  $\mu$  filament wire is compared with a beard hair, and in *fig. 2* in which a grid and a filament wire of a sub-miniature valve are shown in 8-fold magnification. One handful of tungsten ore is sufficient to make a length of about 350 km of 10  $\mu$  wire. Such a length, moreover, satisfies over its entire length the very stringent demands made upon it for the manufacture of radio valves.

*Molybdenum wire* with a thickness varying between 200 and 25  $\mu$  is used for the grids of most miniature and sub-miniature valves. This material is further used as a core or mandrel upon which coiled filaments are wound. At the end of the process the core is removed by dissolving it in a mixture of nitric acid, sulphuric acid and water.

In this article we shall describe the method by which very thin metal wire is manufactured. We

<sup>1)</sup> See K. Rodenhuis, H. Santing and H. J. M. van Tol, The life and reliability of valves, Philips tech. Rev. 18, 193-204, 1956/57.

above this range, there is but little risk of acoustic feedback.

With a view to combating microphony, therefore,  $T/M$  should be made as large as possible. However, the risk of breaking the wire makes it necessary to limit the tension  $T$  to a value considerably below that at which the hot filament would break (e.g. to about one third of the disruptive force).

Because both  $M$  and the permissible value of  $T$  are proportional to the cross-sectional area of the wire, the maximum attainable resonant frequency depends entirely upon the length. In this respect a short filament, and hence a short electrode system, is desirable.

As a tungsten filament can stand a far higher tension than one of nickel, substantially higher resonant frequencies can be obtained with it. Since, however, the internal damping for mechanical vibrations is greater for nickel than for tungsten, the reverberation time is shorter for a nickel wire.

As an example of what can be achieved to avoid microphony it may be mentioned that the resonance frequency of the filament in the sub-miniature valve DF 64 is as high as 8000 c/s.

Finally it should be pointed out that high demands are imposed, not only upon the filaments themselves, but also upon the springs by which they are stretched, in order to keep the wires under an appropriately reproducible tension.

**Summary.** The general endeavour to make radio valves of ever smaller dimensions presents a number of problems, some of which are dealt with in this article.

For the smallest types of so-called sub-miniature valves the old pinch-construction, long since abandoned in the manufacture of normal radio valves, was first employed. More recently sub-miniature valves have been successfully made using a pinch-less construction. For the filaments and grids of these valves extremely thin wires are used. The filament is made of tungsten instead of nickel, which results in a reduction in filament current. By making the electrode system short and by using high-quality springs for stretching the filament, the resonant frequency of the filament can be raised, which gives a decrease in valve microphony.

## THIN TUNGSTEN WIRE FOR SMALL RADIO VALVES

by L. SCHULTINK and P. G. van ZANTEN.

669.27.426:  
621.385.032.213.13:  
621.375.2:621.396.62

*For the manufacture of miniature and sub-miniature valves extremely thin wires of tungsten and other metals are required, as discussed in the preceding article. The article below gives some particulars regarding the manufacture of such thin wire and deals with a number of requirements it has to satisfy.*

The general aim to reduce the dimensions and the power consumption of radio valves has led to the employment of extremely thin metal wire in the construction of filaments, grids and certain other components. Filaments for sub-miniature and miniature valves, of the directly-heated as well as the indirectly-heated type, are made of *tungsten wire* with a thickness of 50 down to 6  $\mu$  — wire of the same material as used for the filaments of incandescent lamps. Certain grids, particularly the "frame grids" used in some valves for professional application<sup>1)</sup>, are also wound of tungsten wire. Furthermore, tungsten wire is used for some of the smaller components of radio valves, such as hooks and tiny springs for stretching the filaments.

An impression of the extreme thickness of such wire is given in *fig. 1* in which a 15  $\mu$  filament wire is compared with a beard hair, and in *fig. 2* in which a grid and a filament wire of a sub-miniature valve are shown in 8-fold magnification. One handful of tungsten ore is sufficient to make a length of about 350 km of 10  $\mu$  wire. Such a length, moreover, satisfies over its entire length the very stringent demands made upon it for the manufacture of radio valves.

*Molybdenum wire* with a thickness varying between 200 and 25  $\mu$  is used for the grids of most miniature and sub-miniature valves. This material is further used as a core or mandrel upon which coiled filaments are wound. At the end of the process the core is removed by dissolving it in a mixture of nitric acid, sulphuric acid and water.

In this article we shall describe the method by which very thin metal wire is manufactured. We

<sup>1)</sup> See K. Rodenhuis, H. Santing and H. J. M. van Tol, The life and reliability of valves, Philips tech. Rev. 18, 193-204, 1956/57.

shall mainly concern ourselves with tungsten wire, since this is made in the smallest diameter, whilst the process requires the highest degree of accuracy. Molybdenum wire, however, can be manufactured by a nearly identical process.

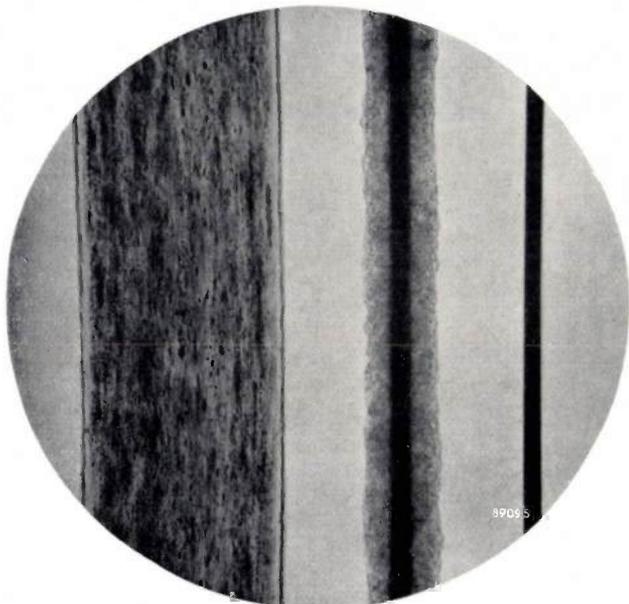


Fig. 1. Tungsten wire of about  $15\ \mu$  thickness (right) compared with a beard hair (left). In the centre a cross-section of the directly-heated cathode of a sub-miniature valve (tungsten wire, coated with an emitting layer). Magnification approx.  $150\times$ . This photograph, and that of fig. 2, were made in the laboratory of Mullard Radio Valve Co., Mitcham, England.

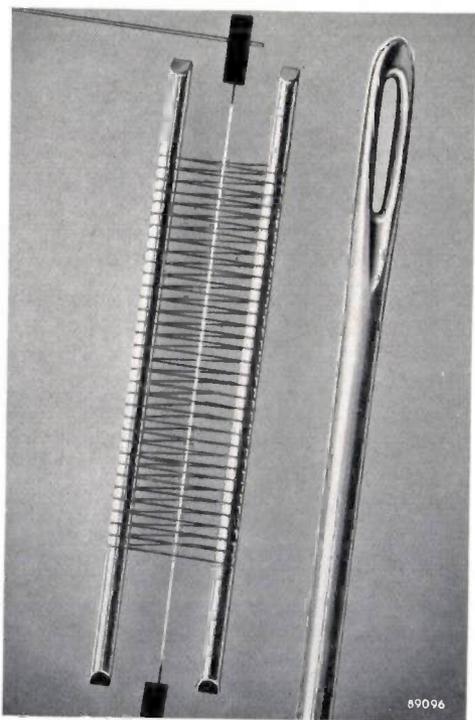


Fig. 2. 8-fold enlargement of a grid and a filament of a sub-miniature valve. The grid is wound of  $25\ \mu$  molybdenum wire; the filament is a tungsten wire of  $11\ \mu$  diameter. The sewing needle gives an impression of the actual size.

#### Material and dimensions of filament wire

The manufacture of filaments for radio valves was initially based on the experience gained in making incandescent lamps. When, between 1930 and 1940, the coiled-coil filament<sup>2)</sup> was introduced and generally adopted, the demands made on the material (tungsten) were to be necessarily far higher than those made on the earlier used single-coil wires.

A coiled-coil filament, when heated between two supports, is permitted only the slightest degree of sagging in view of the far greater risk of short-circuiting between the very close windings. After a thorough investigation into the changes in crystal structure occurring in the filament when the lamp is burning, coiled-coil filaments were successfully made whose sagging was sufficiently small to preclude short-circuiting between the windings. The radio-valve industry was able to profit directly from these results and from the experience of the lamp industry. It was of course additionally necessary, for radio valves, to apply the well-known oxide mixture coating, used as electron-emitting material, to the tungsten filament, the latter acting as the actual heating element.

As the development of the thermionic tube progressed, however, it became increasingly clear that as regards both materials and assembly techniques, the considerations relevant to lamps were not necessarily relevant to valves. The special requirements of valves had to be taken into account. For example, in receiver valves of the directly heated type, tungsten as cathode material was superseded by nickel, the latter providing a better base for the emitting substance.

As mentioned in the preceding article<sup>3)</sup>, there has been a return to tungsten for certain types of valves, particularly for sub-miniature valves. The main reason for this was that for a further reduction of the filament current ever thinner filaments had to be used, for which tungsten is more suitable than nickel. The extremely low current consumption of the sub-miniature valves is indeed mainly attributable to the possibility of manufacturing such thin tungsten wires.

The use of thin tungsten wire, however, is not restricted to sub-miniature valves alone. For valves requiring a larger filament current, it is advantageous to use a cathode consisting of more than one thin wire in parallel. The slope of a valve of this con-

<sup>2)</sup> See W. Geiss, The development of the coiled-coil lamp, Philips tech. Rev. 1, 97-101, 1936.

<sup>3)</sup> B. A. Cant, Problems in the construction of small radio valves, Philips tech. Rev. 18, 217-222, 1956/57 (No. 8).

struction, for a given anode current, is greater than that of a valve passing the same filament current through a thicker, single wire.

If the cathode diameter is small with respect to the anode diameter, the former has virtually no effect on the slope at a given current<sup>4</sup>). It is further found that in a cathode consisting of a number of thin wires in parallel, each of these wires contributes about as much to the slope as the amount that can be obtained with one of these wires separately. Hence, by building up the cathode from a number of thin wires the total slope is increased.

Examples of multi-wire filaments include the output valve DL 96, with a cathode of two parallel tungsten wires of 11  $\mu$  diameter; the filament current is 50 mA, whilst the max. emission current is 6 mA. The output valve DL 41 employs a cathode of four parallel-tungsten wires, likewise of 11  $\mu$  diameter; the filament current of this tube is 100 mA, whilst the max. emission current amounts to 12 mA.

#### The manufacture of very thin tungsten wires

For wire with a diameter greater than 10  $\mu$  the normal drawing process, by means of diamond dies, presents no major difficulties<sup>5</sup>).

Drawing still thinner wire, however, becomes increasingly difficult. In spite of the considerable tensile strength of tungsten, a pull of only 25-30 gm will break a 10  $\mu$  wire at room temperature; for a diameter of 7.5  $\mu$  the breaking stress is reduced to about 15 gm, and for a diameter of 5  $\mu$  to about 5 gm. Any mechanical working of such wire presents, of course, enormous difficulties.

For the drawing of this wire, a whole series of dies is required, each with a diameter differing only very slightly from the previous one. It is extremely difficult to drill these very small apertures in the diamonds to the narrow tolerances required. The introduction of "electric drilling" (various electro-sparking techniques)<sup>6</sup>) has admittedly simplified the production of diamond dies with very small apertures, but in spite of the advantages of this process it is still far from simple to meet the requirements specified above.

The most commonly used method of making very thin wire is by drawing it to the smallest diameter at which it can still be handled and then further

reducing the diameter by an *etching* process. The etching can be done chemically or electrolytically. In the former case either a gas can be used, e.g. oxygen, water vapour or chlorine, or a molten oxidizing agent, such as a mixture of potassium nitrite and potassium nitrate. An aqueous solution of an oxidizing agent such as a mixture of hydrofluoric and nitric acid or of potassium ferricyanide (red prussiate of potash) and sodium hydroxide can also be employed.

A great drawback of all chemical etching methods is the difficulty of carrying out the process with sufficient uniformity. Much simpler in this respect is the process of *electrolytic etching*<sup>7</sup>), in which the wire is passed through a dilute alkaline solution, whilst an electric potential is applied between the wire and an electrode placed in the liquid. If the wire is connected to the positive terminal of the voltage source, its diameter will diminish. An attractive feature of this process is that it can be readily carried out continuously and at room temperature. Furthermore, the surface of the wire is thoroughly cleaned at the same time, this being essential for good emission of the cathode made of this wire. It is for this last reason that wires thicker than 10  $\mu$ , which could be fabricated by drawing alone, are also subjected to a subsequent etching process.

The use of the etching process imposes severe demands upon the thicker wire which is to be so treated. Its diameter should be as uniform as possible, since any deviations from the nominal value will become relatively greater after etching. In view of the low tensile strength of the wire, the whole etching apparatus must be constructed with the utmost care. The bobbin from which the wire is reeled off and the guide pulleys should run with a minimum of friction. *Fig. 3* shows an installation for electrolytic etching. The wire successively passes several containers with the etching fluid.

#### Properties required in the finished wire

To be suitable for the filaments for radio valves, the wire has to satisfy various specific requirements. Some of these requirements can be conveniently expressed as physical quantities, such as diameter uniformity, weight and resistance per unit length, tensile strength and temperature coefficient of resistance. We shall first deal with these requirements, and then mention a few that are less readily expressed as physical quantities.

<sup>4</sup>) See e.g. H. Rothe and W. Kleen, *Grundlagen und Kennlinien der Elektronenröhren*, Akademische Verlagsgesellschaft Becker & Erler Kom.-Ges., Leipzig, 1943, 2nd ed., chapters 20 and 4.

<sup>5</sup>) See e.g. J. D. Fast, *The preparation of metals in a compact form by pressing and sintering*, Philips tech. Rev. 4, 309-316, 1939.

<sup>6</sup>) C. G. Peters, W. B. Emerson, K. N. Neffen, F. K. Harris and J. L. Cooper, *Electrical methods for diamond-die production*, Industrial Diamond Review 8, 44, 1948.

<sup>7</sup>) See J. O'M. Bockris and D. F. Parsons, *Preparation of thin tungsten fibres and their use in place of quartz*, J. sci. Instr. 30, 340, 1953.

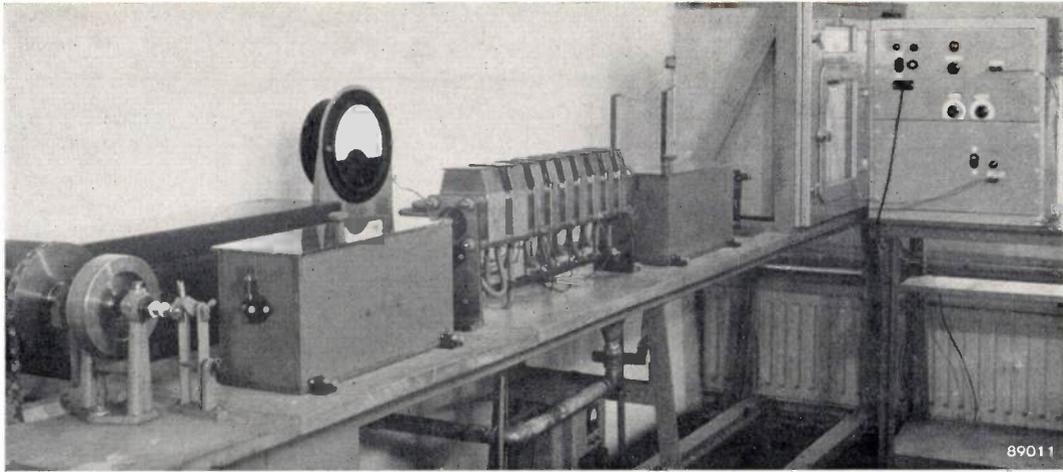


Fig. 3. Installation for the electrolytic etching of very thin metal wire. The wire passes successively through a number of baths of etching fluid.

The diameter of the very thin wires considered here can no longer be accurately gauged with the aid of a microscope. In order to test the uniformity of the diameter, moreover, so many measurements must be made that this method is unsuitable for a continuous checking during the manufacturing process. Several continuous methods for gauging the diameter during processing have been described in the literature. Evans, Marriner and Morgan<sup>8)</sup> describe a pneumatic method of gauging. The wire is passed, via two diamond apertures, through a closed chamber into which a gas is fed at a certain constant pressure. If the wire is perfectly uniform in diameter, a constant flow of gas escapes via the apertures in the diamonds. If the diameter is not uniform, on the other hand, the flow of escaping gas will vary, causing fluctuations of the pressure in the compartment. In another gauging method, the wire is passed through a capacitor, variations in diameter thus gauging variations in the capacitance<sup>9)</sup>. In either case deviations from the average diameter can be recorded and, by means of special measures, deviations from roundness of the wire can also be detected.

The diameter of the wire may also be evaluated from the *weight per unit length*. In practice this is always done by weighing a length of wire of 200 mm with a sensitive torsion balance. In this way, obviously, only the average diameter of the length of wire can be evaluated.

<sup>8)</sup> J. C. Evans, R. S. Marriner and J. G. Morgan, Continuous pneumatic gauging of fine wire, *The Machinist* **96**, 679-682, 1952.

<sup>9)</sup> W. W. Loebe and C. Samson, Beobachtungen und Registrierungen von Durchmesserschwankungen dünner Drähte, *Z. techn. Phys.* **9**, 414-419, 1928.

To measure the *electrical resistance* per unit of length during the manufacturing process, the wire is passed through two mercury contacts, the wire between these contacts forming one arm of a Wheatstone bridge<sup>10)</sup>. The sensitivity of these measurements can be increased by reducing the distance between the contacts and the speed of the wire. As an example *fig. 4* shows the result of a measurement of this kind for a tungsten wire of  $11 \mu$ , passing at a speed of 15 m/min. between contacts separated by a distance of 20 cm. It is found here that the maximum deviations from the average value amount to  $\pm 0.3\%$ . Wires showing variations in the resistance no greater than this are quite suitable for the filaments of radio valves for battery supply.

The *tensile strength* of the wire can be readily determined by measuring the pull under which a wire breaks. Since this is a destructive test, this value can only be established by sample testing, and not

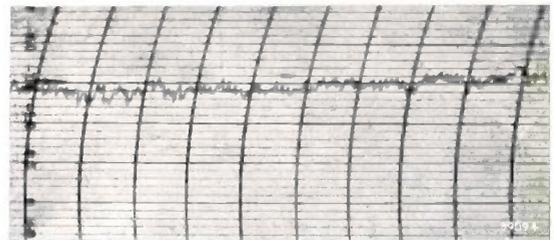


Fig. 4. Recording of the resistance per unit length of a tungsten wire of  $11 \mu$ , passing at a speed of 15 m/min through the arm of a bridge circuit. The distance between the contacts was 20 cm. The maximum deviation from the average value is  $\pm 0.3\%$ .

<sup>10)</sup> See K. Dahl and J. Kern, Ein schreibendes Messgerät zur Messung der Querschnittschwankungen feiner Drähte, *E.T.Z.* **57**, 1423-1425, 1936.

continuously during the manufacturing process. The same applies to the *temperature coefficient of resistance*. Here, too, samples are taken and tested by measuring the resistance at various temperatures in a Wheatstone bridge.

Among the requirements that are difficult to define or measure as physical quantities are creep, brittleness, roundness, "straightness", surface texture and the presence of splits.

*Creep* is the slow continuous elongation of the wire when subjected to a constant tensile force. While the creep may be negligible at room temperatures, it can be considerable at the high temperatures at which a valve filament operates. As a result the stress in the filament will be reduced, this increasing the risk of microphony. If the tensile load is large, the wire may even break in the long run. In order to obtain data of practical significance, creep should be measured at high temperature. This measurement requires high-precision equipment and is, moreover, very time-consuming. For these reasons creep measurements as such are as a rule dispensed with, although information of practical value in this connection is often obtained as follows. Samples of the wire are taken and, by means of a small weight or a spring, each wire is subjected to the same tensile stress, which is a multiple of the normal filament stress. The wires are heated to about 800 °C by means of an electric current. If the time that elapses until the various wires break is noted, a curve may be plotted of the number of unbroken wires as a function of the time. The form of this "survival curve" varies according to the

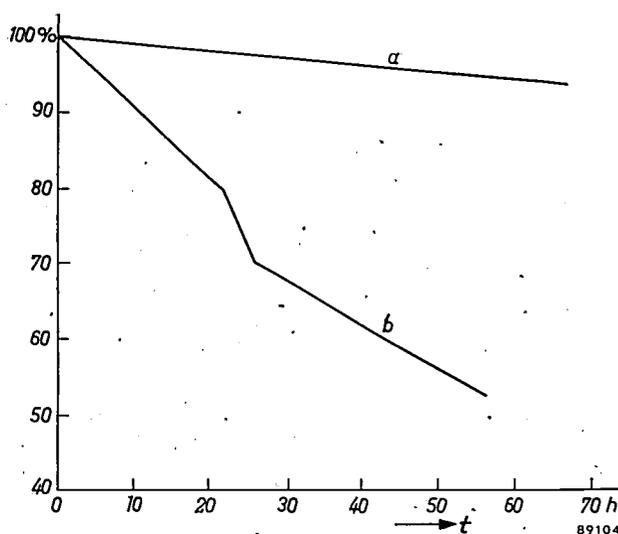


Fig. 5. "Survival curves", indicating the percentage of unbroken wires as a function of the time  $t$  in hours. The wires were heated in vacuo up to a temperature of 775 °C and subjected to a tensile stress which is a multiple of the normal filament stress. Curve  $a$  represents a good quality wire; curve  $b$  a wire of poor quality.

process and conditions of manufacture of the wire. Two such curves are shown in fig. 5. With the aid of the survival curve of a given batch of wire, the valve designer can decide what voltage and tension may be applied to the wire while keeping the risk of broken filaments in the course of the normal life of the valve below an acceptable limit.

The *brittleness* is generally assessed by bending the wire through a given angle over an edge of a certain radius and counting the number of times this can be repeated before the wire breaks.

For judging the *roundness*, it is possible, but rather difficult to detect slight discrepancies in the circular shape of the cross-section of the final product. A preventive check on the aperture of the last die through which the wire is drawn is as a rule more effective.

As a check on the presence of irregularly distributed internal stresses, a given length of wire is often freely suspended and the extent to which the wire deviates from a straight line is observed. For the sake of brevity this is often referred to as the "straightness" of the wire.

The *surface texture* of the wire can be varied considerably by etching. The surface may vary from glossy to dull. As an example fig. 6a and b show the surfaces of two wires etched by different processes, at the same magnification.

The detection of local cracks in the wire is extremely difficult. Even with the aid of a microscope it is very difficult to determine whether a wire is cracked or merely shows a deep groove. The issue may be decided by examining a lateral cross-section. As a rule cracks occur when the wire is also in other respects in a generally poor condition.

#### Coating the wire with alkaline earth carbonates

For use as cathodes in radio valves the wire has to be coated with the carbonates of barium, strontium and sometimes calcium. There are two methods of coating, viz.:

1. *Continuous immersion* ("drag coating"). The wire is passed several times through a suspension of the alkaline earth carbonates to which is added a certain amount of the nitrate of one or more of the above-mentioned metals. Each time the wire is passed through the suspension, a thin layer adheres the wire, which is subsequently sintered to the surface by heating. The nitrates serve here as an inorganic binder. The layer thickness obtained depends upon the speed of the wire, the viscosity of the suspension and the number of times the process is repeated.

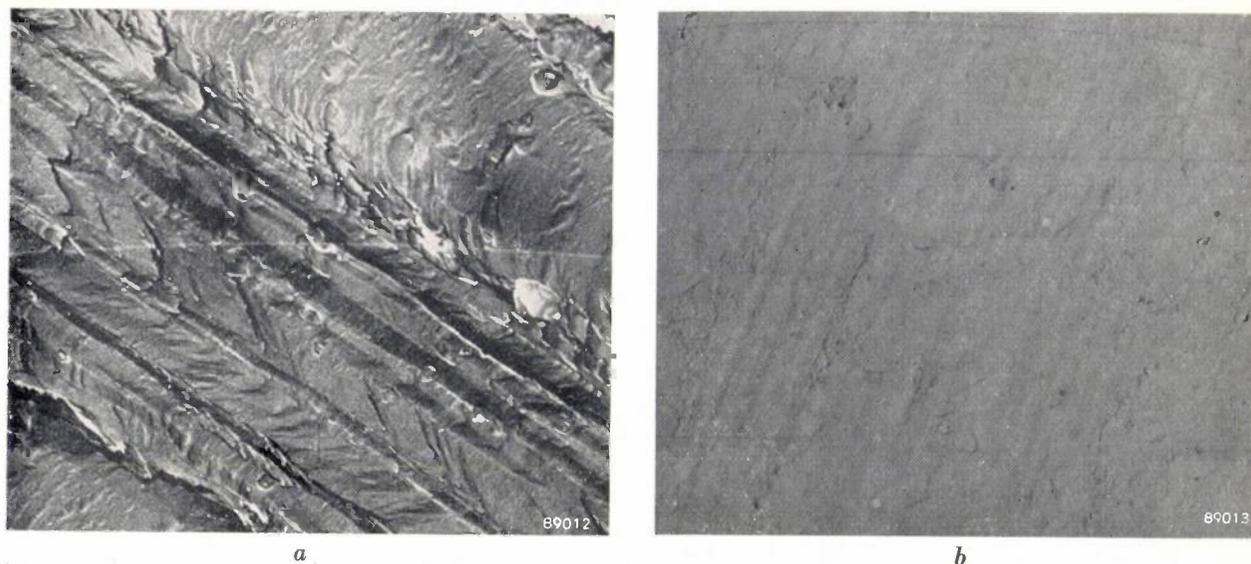


Fig. 6. The etching process has a profound influence on the surface texture of the wire. The figures show a roughly etched wire (a) and a smoothly etched wire (b). Magnification 16 000  $\times$ .

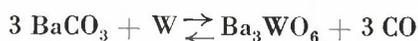
2. *Electrophoresis*<sup>11)</sup>. The wire is passed through a bath containing a suspension of charged carbonate particles. Under the influence of a voltage of some tens of volts, applied between the wire and an anode in the suspension, the carbonate particles migrate towards the wire, to which they adhere. Appropriate selection of voltage and passing speed of the wire makes it possible to obtain the desired layer thickness in one operation. After drying, an additional coating of a binder is deposited by passing the wire through a solution of this binder.

#### Decomposition of the carbonates

When the radio valve is being exhausted the wire must be heated in order to convert the carbonate into oxides. At the same time any organic binder present is carbonized and burnt ( $\text{CO}_2 + \text{C} \rightarrow 2\text{CO}$ ).

When tungsten is heated with barium carbonate or barium oxide, certain reactions occur with the metallic tungsten; these have been investigated in the Philips Laboratory of Irvington, U.S.A.<sup>12)</sup>

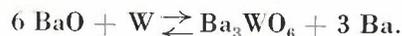
At a temperature of approx. 600 °C, the reaction



is initiated. Since the formation of the semi-conducting barium tungstate is undesirable, this reaction should be avoided. If the temperature is quickly raised to 800-900°, the barium carbonate decomposes according to:



If, however, the temperature is raised above 1000 °C, the barium oxide is reduced by the tungsten with the formation of barium tungstate as well as free barium, according to:



In this way the free barium required for the proper functioning of an oxide cathode is obtained. In view of the fact that only a small quantity of free barium is required, whilst at the same time the formation of the semi-conducting barium tungstate is undesirable<sup>13)</sup>, heating at high temperature should take place for a short period only.

The last reaction probably also occurs, if only on a very small scale, during the normal functioning of a radio valve; the free barium lost through evaporation or combination with residual gas is thus made up for.

On normal indirectly-heated oxide-coated cathodes, a carbonate layer with a thickness of 75  $\mu$

<sup>11)</sup> See S. A. Troelstra, Applying coatings by electrophoresis, Philips tech. Rev. 12, 293-303, 1950/51; H. C. Hamaker, A general theory of lyophobic colloids, Rec. Trav. chim. Pays-Bas 55, 1015-1026, 1936 and 56, 3-25, 1937; H. C. Hamaker, A system of colloid phenomena, Rec. Trav. chim. Pays-Bas 56, 727-747, 1937; L. E. Grey and R. O. Jenkins, Electrophoresis in the valve industry, Electronic Engineering 26, 402-405, 1954.

<sup>12)</sup> R. C. Hughes, P. P. Coppola and H. T. Evans, Chemical reactions in barium oxide and tungsten emitters, J. appl. Phys. 23, 635-641, 1952.

<sup>13)</sup> Similar semi-conducting products, formed during the reduction of barium oxide have recently been the subject of further research; see e.g. J. F. Waymouth, Deterioration of oxide-coated cathodes under low-duty-factor operation, J. appl. Phys. 22, 80-86, 1951 and L. S. Nergaard and R. M. Matheson, Studies of the interface layer in oxide cathodes, R. C. A. Review 15, 335-361, 1954.

is usually deposited. For a directly-heated cathode with a diameter of only 8 or 11  $\mu$  such a layer is obviously far too thick. The usual layer thickness on a tungsten wire of 11  $\mu$  is 20  $\mu$ , and even here the ratio of the diameter of the coated wire to that of the tungsten core is so great that the temperature of the core differs a great deal from that at the surface. It is owing to this temperature difference, however, that thin cathodes with a tungsten base can retain a better activity than thicker cathodes. As tungsten is a rather poor reducing agent for the alkaline earth oxides (as compared with silicon, magnesium or aluminium, which are alloyed in normal cathode nickel), a high temperature of the core is required to ensure a permanent supply of free barium. The temperature at the surface, however, determines the rate at which the barium disappears through evaporation and combination with residual gas, and here a not-too-high temperature is favourable.

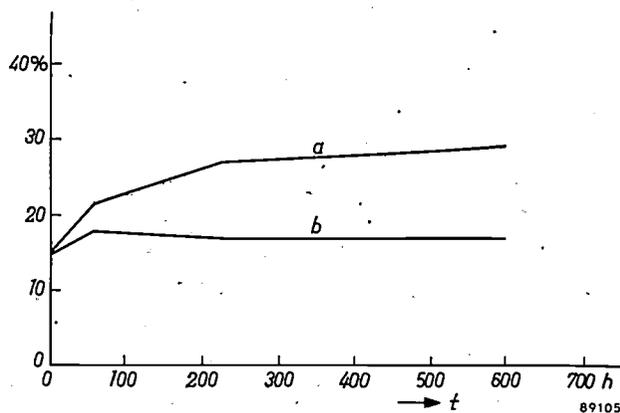
It is interesting to note that for a given filament current the temperature of the core depends very little on the thickness of the coating, unlike the temperature at the surface which is strongly correlated with the thickness.

#### Results of life tests

The emission current that can be supplied by an oxide-coated cathode with a core of thin tungsten wire, is difficult to predict with any accuracy, since this depends greatly on the purity of the materials involved. The quantity of oxide on a length of wire of 2 cm having a diameter of 11  $\mu$  amounts to approx. 30  $\mu\text{g}$ . If we assume that 0.01% of this consists of free barium<sup>14</sup>), then it follows that 0.003  $\mu\text{g}$  of active material is available on the cathode. Clearly, the production of valves in which the quantity of free barium remains roughly constant during the entire working life, whilst at the same time the formation of undesirable tungstates is kept within reasonable limits calls for the strictest precautions and the highest precision. A further requirement is that the filament voltage should be permitted to

<sup>14</sup>) R. O. Jenkins and R. H. C. Newton, Free barium in and evaporated from oxide cathodes, *Nature* 163; 572, 1949.

vary within the wide limits between which the terminal voltages of dry batteries are apt to vary. Nevertheless it is found possible to obtain satisfactory results in practice. As an example, *fig. 7* shows



*Fig. 7.* Results of life tests on a number of DL 96 valves. Curve *a* was measured on valves under a filament voltage of 1.5 V, having a cathode-emission current of 5.1 mA during the life test. Plotted along the ordinate is the per cent change in slope when the filament voltage is varied from 1.4 to 1.0 V (at an anode current of 5 mA). Curve *b* relates to valves under a filament voltage of 1.1 V, having an emission current of 4 mA throughout the life test. The time  $t$ , along the abscissa, is plotted in hours.

some results of life test on the miniature valve DL 96. The per cent change in slope, measured with varying filament voltages, as plotted in this diagram, appears to lie well within acceptable limits after several hundreds of burning hours.

**Summary.** Filaments for electronic valves were initially made of tungsten. Later this material was superseded by nickel, but in the sub-miniature valves developed in the last two decades tungsten has been reinstated as filament material. The reason for this is the greater tensile strength of this material, thanks to which the filament can be made thinner, whilst a far smaller filament current will suffice. The most commonly employed method for making very thin wire is the electrolytic etching of thicker wire. The principal requirements for the wire to be suitable as filament material are high tensile strength and low creep. The wire is coated with a layer of alkaline earth carbonates. The principal method of coating is by means of electrophoresis. During pumping of the valve, prior to sealing, the carbonates are decomposed into oxides by heating the filament. For very thin filament wires the ratio of the diameter of the coated wire to that of the bare core is fairly high. Owing to the difference in temperature thus caused between core and surface, such cathodes will retain their activity far longer.

## NUCLEAR X-RAY SOURCES

by S. FINE \*) and C. F. HENDEE \*).

539.167.3:537.531:621.387.4

*With the rapidly expanding availability of artificially produced radioisotopes (radioactive nuclides) it has become possible to use many of these as radiation sources for routine work. Nuclides directly emitting soft X-rays have been put to such use in recent years. The authors have made a survey of X-ray emitting nuclides and have prepared a number of X-ray sources of this type ("nuclear X-ray sources") which have proved very valuable in work on and with X-ray counter tubes.*

Radioactive substances, either natural or artificial, have been employed for a long time as sources of  $\alpha$ - and  $\beta$ -particles and of high energy photons ( $\gamma$ -rays). A classical example is natural radioactive polonium, specimens of which are often used for demonstrating the effects of  $\alpha$ -radiation or for checking apparatus sensitive to it. A younger member of this family is the artificial radioisotope of cobalt (Co 60) emitting 1.17 MeV and 1.33 MeV  $\gamma$ -rays. Radiation of similar energy, which is used, inter alia, for industrial radiography and for therapeutic irradiation, may also be obtained from X-ray tubes run at voltages up to 2 MeV, or from betatrons or linear accelerators<sup>1)</sup>. Nevertheless the radioactive  $\gamma$ -ray source has proved a very valuable extension of the array of high-energy X-ray machines.

In addition to the energetic particles and high energy photons, resulting from the decay of radioactive nuclei, photons of lower energy, e.g. less than 100 keV (soft X-rays), are in some cases emitted. Because the investigation of nuclear processes usually involves an interpretation of high energy events, there has been little emphasis on the low energy phenomena. In fact the latter often appear only as perturbations or corrections in nuclear theory (background or noise).

Only in recent years has it been realized that the emission of low energy X-rays by some radioactive nuclides can be put to good use<sup>2)</sup>, just as useful applications have been found for  $\gamma$ -ray emitters such as cobalt 60. Interest in this possibility was prompted by the development of X-ray quantum counting techniques. For checking and calibrating

proportional and scintillation counters developed for X-ray detection<sup>3)</sup>, very stable monochromatic X-ray sources were required. X-ray emitting nuclides ("nuclear X-ray sources"), while lacking the easy controllability and the very large radiation output of modern X-ray generating equipment, usually greatly surpass the latter in stability and therefore promised to be useful in work with X-ray counters.

Unfortunately, nuclear X-ray sources have not yet become a commercially available product, probably owing to the tendency among physicists to be more preoccupied with high energy phenomena. It seemed desirable, therefore, to survey the presently known artificially produced radioisotopes from the viewpoint of X-ray production. This survey is offered in this article, together with a discussion of the underlying phenomena in the nuclear X-ray source and a description of some sources which have been developed and are currently used in the Philips Laboratories at Irvington<sup>4)</sup>.

### Mechanisms of X-ray emission by radioactive nuclides

There are two decay processes of radioactive nuclei which lead directly to the emission of X-rays<sup>5)</sup>:

- 1) Internal conversion of emitted gamma rays.
- 2) Capture of orbital electrons.

<sup>3)</sup> See e.g. W. Parrish and T. R. Kohler, *Rev. sci. Instr.* **27**, 795-808, 1956; also P. H. Dowling, C. F. Hendee, T. R. Kohler and W. Parrish, *Counter tubes for X-ray analysis*, Philips tech. Rev., to appear shortly.

<sup>4)</sup> Part of the material of this article was presented by S. Fine and C. F. Hendee at the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, Feb. 1956.

<sup>5)</sup> X-rays can also be produced indirectly, by  $\alpha$ ,  $\beta$  or  $\gamma$ -rays emitted from a radionuclide and impinging on a suitable target material. These methods have been known and applied for a long time. I. Curie and F. Joliot discussed this possibility with reference to  $\alpha$ -rays (*J. Phys. Rad.* **2**, 20-28, 1931); for  $\beta$ -ray excitation, see L. Reiffel, *Nucleonics* **13**, No. 3, March 1955, pp. 22-24. Although these methods certainly should not be overlooked, we shall not consider them in this article.

\*) Philips Laboratories, Irvington-on-Hudson, N.Y., U.S.A.

<sup>1)</sup> See e.g. I. Kaplan, *Nuclear physics*, Addison-Wesley, Cambridge (Mass.), 1955. See also Philips techn. Rev. **4**, 153-161, 1939; **15**, 1-26, 1953/54; **17**, 31-33, 1955/56.

<sup>2)</sup> G. M. Insch, *Phil. Mag.* **7**, **41**, 857-862, 1950.

P. Rothwell and D. West, *Proc. Phys. Soc.* **A63**, 541-543, 1950.

W. V. Mayneord, *Brit. J. Radiology* **25**, 517-525, 1952.

Both these mechanisms can be readily described when it is remembered how the characteristic X-ray emission of an atom, say the K-lines, is produced: an electron is removed from an orbit in the K-shell, and the vacancy so produced is filled by an electron from one of the outer shells, each of the possible transitions being accompanied by the emission of the energy difference as an X-ray quantum of corresponding wavelength. In the target of an X-ray tube, the removal of a K-electron (or of an L-electron for the emission of L-lines, etc.) is effected by electron impact. This is different from the above-mentioned mechanisms occurring in radioactive nuclides.

In the first mechanism, a nucleus being in an excited state decays to its normal state and emits a gamma quantum; the latter has a finite probability of ejecting an orbital electron of the atom. The efficiency of this process (coefficient of "internal conversion", i.e. average number of electrons ejected per gamma quantum) increases with the atomic number  $Z$  and with the lifetime of the excited state, and rather rapidly with decreasing gamma quantum energy. The efficiency will be high and may even approach unity for nuclei in a *metastable* ("isomeric") excited state, since such states are long-lived and generally have low excitation energies (small gamma quanta). Nuclei of this type can therefore be effective X-ray sources (the effectiveness, however, depending also on other factors, see hereafter).

In the second mechanism, the orbital electron is not ejected but swallowed (captured) by the unstable nucleus itself. In this process the nucleus of atomic number  $Z$  loses one positive charge, and the X-ray photons subsequently emitted by the filling of the orbital vacancy are characteristic of the new atom, with atomic number  $Z-1$ . This should be contrasted with the first mechanism, in which no change of nuclear charge takes place and the characteristic X-ray spectrum of the atom of atomic number  $Z$  is emitted.

It should be noted that in both cases a pure characteristic X-ray spectrum is emitted, *not* containing a Bremsstrahlung continuum which is unavoidable in X-ray tubes. This is an important feature for X-ray sources whose emissions are used as wavelength standards, e.g. in high resolution spectroscopy.

The K-electrons are most frequently ejected or captured by the nuclear processes described above, because of their close proximity to the nucleus. However, there is a finite probability for the ejection or capture of L-, M- etc. electrons.

### List of X-ray emitting nuclides

About 100 radioactive nuclides decaying by the above processes and producing X-rays are known, with half-lives varying from hours to thousands of years and X-ray wavelengths varying from 2.5 to 0.15 Å (energies 5 to 80 keV). Table I is a partial listing of these nuclides together with some of their important characteristics. The information was assembled from literature and official publications on isotopes available from atomic piles or cyclotrons<sup>6)</sup>. In assembling the list, an attempt was made to select nuclides with K emissions in a somewhat continuous series up to 75 keV. Nuclides with a half-life of less than ten days are not listed, since their radiation intensity will decrease too rapidly for most practical purposes.

The nuclides are produced in nuclear reactions caused by the bombardment of target substances. The reactions are listed in the third column of the table. In all cases the target materials are naturally stable isotopes, whose abundance is also given in the table; the increased abundance indicated in the next column refers to enriched specimens obtainable from the Oak Ridge National Laboratories, U.S.A. The mechanism of X-ray emission is indicated under the heading "Decay mode". In the next column the table gives the element whose characteristic K-lines are emitted, and the energy of  $K\alpha_1$ <sup>7)</sup>. In most cases other emissions, such as  $\gamma$ -radiation, conversion electrons ( $e^-$ ) or electrons from the nucleus ( $\beta^-$ ) accompany X-ray emission (last column). These need not be an interfering factor in many applications. If necessary, the electrons may be eliminated by metal foil or plastic film filters, while the  $\gamma$ -rays may be minimized by electronic discrimination<sup>3)</sup>.

It is to be noted that Table I includes nuclides that emit the characteristic radiations of the X-ray tube target materials commonly used for X-ray diffraction and spectrochemical analysis, namely Cr, Fe, Co, Cu, Mo, Ag and W.

In some cases the new nucleus resulting from an electron capture process is in a metastable excited

<sup>6)</sup> J. M. Hollander, J. Perlman and G. T. Seaborg, Table of isotopes, Rev. mod. Phys. 25, 469-651, 1953. A. H. W. Aten and J. Halberstadt, The production of radioisotopes, Philips tech. Rev. 16, 1-12, 1954/55. This article also gives a brief exposition of a number of nuclear reactions and of chemical methods used in the separation of the isotopes. J. R. Bradford, Chart of the isotopes, publ. Harshaw Scientific. G. Friedländer and M. Perlman, Chart of the nuclides (revised by J. R. Stehn), publ. General Electric, U.S.A.  
<sup>7)</sup> S. Fine and C. F. Hendee, X-ray critical-absorption and emission energies in keV, Nucleonics 13, No. 3, March 1955, pp. 36-37.

Table I. Partial list of X-ray emitting radioactive nuclides, with their half-life, nuclear reactions by which they are produced, decay mode, and emitted radiation. Under decay mode, EC stands for electron capture process, IT for internal conversion or transition and m for metastable nucleus; competing processes by which some nuclei can decay ( $\beta^+$ -emission) are also indicated; in some cases decay involves successive processes.

Radioactive nuclide		Target			Radiation		
Symbol and mass number	Half-life	Reaction	Natural abundance %	Increased abundance %	Decay mode	X-ray $K\alpha_1$ (keV)	Other
Be <sup>7</sup>	53 d	Li <sup>7</sup> (d,2n)Be <sup>7</sup>	92.6	—	EC	Li 0.052	$\gamma$
Ca <sup>41</sup>	1.1 × 10 <sup>5</sup> y	Ca <sup>40</sup> (n, $\gamma$ )Ca <sup>41</sup>	96.97	—	EC	K 3.31	—
V <sup>49</sup>	1 y	Ti <sup>48</sup> (d,n)V <sup>49</sup>	73.45	99.2	EC	Ti 4.51	—
Cr <sup>51</sup>	27 d	Cr <sup>50</sup> (n, $\gamma$ )Cr <sup>51</sup>	4.31	88.3	EC	V 4.95	$\gamma$
Mn <sup>54</sup>	300 d	{ Cr <sup>53</sup> (d,n)Mn <sup>54</sup> Fe <sup>50</sup> (d, $\alpha$ )Mn <sup>54</sup>	{ 9.55 91.68	{ 92.1 99.9	{ EC	{ Cr 5.41	{ $\gamma$
Fe <sup>55</sup>	2.94 y	Fe <sup>54</sup> (n, $\gamma$ )Fe <sup>55</sup>	5.84	93.3	EC	Mn 5.90	—
Co <sup>57</sup>	270 d	Fe <sup>56</sup> (d,n)Co <sup>57</sup>	91.68	99.9	$\beta^+ \rightarrow$ IT(Fe <sup>57m</sup> )	Fe 6.40	$\gamma, e^-$
Co <sup>58</sup>	72 d	Fe <sup>57</sup> (d,n)Co <sup>58</sup>	2.17	87.3	EC, $\beta^+$	Fe 6.40	$\gamma$
Ni <sup>59</sup>	7.5 × 10 <sup>5</sup> y	Ni <sup>58</sup> (n, $\gamma$ )Ni <sup>59</sup>	67.76	99.3	EC	Co 6.93	—
Zn <sup>65</sup>	250 d	{ Zn <sup>64</sup> (n, $\gamma$ )Zn <sup>65</sup> Cu <sup>65</sup> (d,2n)Zn <sup>65</sup>	{ 48.89 30.9	{ 93.4 98.2	{ EC, $\beta^+$	{ Cu 8.05	{ $\gamma$
Ge <sup>71</sup>	11.4 d	Ge <sup>70</sup> (n, $\gamma$ )Ge <sup>71</sup>	20.55	88.1	EC	Ga 9.25	—
As <sup>73</sup>	76 d	Ge <sup>72</sup> (d,n)As <sup>73</sup>	27.37	89.2	EC	Ge 9.89	$\gamma, e^-$
Se <sup>75</sup>	127 d	Se <sup>74</sup> (n, $\gamma$ )Se <sup>75</sup>	0.87	33.1	EC	As 10.54	$\gamma, e^-$
Y <sup>88</sup>	105 d	Sr <sup>88</sup> (d,2n)Y <sup>88</sup>	82.56	99.7	EC, $\beta^+$	Sr 14.16	$\gamma$
Mo <sup>93</sup>	>2 y	Mo <sup>92</sup> (n, $\gamma$ )Mo <sup>93</sup>	15.86	95.5	EC	Nb 16.61	—
Tc <sup>95</sup>	60 d	Mo <sup>94</sup> (d,n)Tc <sup>95m</sup>	9.12	79.1	EC	Mo 17.48	$\gamma$
Pd <sup>103</sup>	17 d	{ Rh <sup>103</sup> (d,2n)Pd <sup>103</sup> Pd <sup>102</sup> (n, $\gamma$ )Pd <sup>103</sup>	{ 100.00 0.8	{ — —	{ EC $\rightarrow$ IT(Rh <sup>103m</sup> )	{ Rh 20.21	{ $\gamma, e^-$
Cd <sup>109</sup>	330 d	{ Cd <sup>108</sup> (n, $\gamma$ )Cd <sup>109</sup> Ag <sup>109</sup> (d,2n)Cd <sup>109</sup>	{ 0.89 48.65	{ 24.8 99.6	{ EC $\rightarrow$ IT(Ag <sup>109m</sup> )	{ Ag 22.16	{ $\gamma, e^-$
Sn <sup>113</sup>	112 d	Sn <sup>112</sup> (n, $\gamma$ )Sn <sup>113</sup>	0.95	72.5	EC	In 24.21	$\gamma$
In <sup>114</sup>	49 d	{ In <sup>113</sup> (n, $\gamma$ )In <sup>114m</sup> Cd <sup>114</sup> (d,2n)In <sup>114m</sup>	{ 4.23 28.86	{ 65.4 94.2	{ IT	{ In 24.21	{ $\gamma, \beta^-, e^-$
Ba <sup>133</sup>	9.5 y	Cs <sup>133</sup> (d,2n)Ba <sup>133m</sup>	100.00	—	IT $\rightarrow$ EC(Ba <sup>133</sup> )	Cs 30.97	$\gamma, e^-$
Cs <sup>137</sup>	33 y	Fission product	—	—	$\beta^- \rightarrow$ IT(Ba <sup>137m</sup> )	Ba 32.19	$\gamma, e^-$
Sm <sup>145</sup>	410 d	Sm <sup>144</sup> (n, $\gamma$ )Sm <sup>145</sup>	3.16	—	{ EC 30y EC (Pm <sup>145</sup> )	{ Pm 38.65 Nd 37.36	{ $\gamma, e^-$
Gd <sup>153</sup>	236 d	{ Gd <sup>152</sup> (n, $\gamma$ )Gd <sup>153</sup> Eu <sup>153</sup> (d,2n)Gd <sup>153</sup>	{ 0.2 52.23	{ — —	{ EC	{ Eu 41.53	{ $\gamma, e^-$
Dy <sup>159</sup>	140 d	Tb <sup>159</sup> (d,2n)Dy <sup>159</sup>	100.00	—	EC	Tb 44.47	—
Yb <sup>169</sup>	33 d	Tm <sup>169</sup> (d,2n)Yb <sup>169</sup>	100.00	—	EC	Tm 50.73	$\gamma, e^-$
Tm <sup>170</sup>	127 d	Tm <sup>169</sup> (n, $\gamma$ )Tm <sup>170</sup>	100.00	—	$\beta^- \rightarrow$ IT(Yb <sup>170m</sup> )	Yb 52.36	$\beta^-, \gamma, e^-$
Hf <sup>175</sup>	70 d	Lu <sup>175</sup> (d,2n)Hf <sup>175</sup>	97.40	—	EC	Lu 54.06	$\gamma, e^-$
W <sup>181</sup>	140 d	Ta <sup>181</sup> (d,2n)W <sup>181</sup>	100.00	—	EC	Ta 57.52	$\gamma, e^-$
Re <sup>183</sup>	120 d	Ta <sup>181</sup> ( $\alpha$ ,2n)Re <sup>183</sup>	100.00	—	EC	W 59.31	$\gamma, e^-$
Re <sup>184</sup>	50 d	Re <sup>185</sup> (n,2n)Re <sup>184</sup>	37.07	85.4	EC	W 59.31	$\gamma, e^-$
Os <sup>185</sup>	97 d	Re <sup>185</sup> (d,2n)Os <sup>185</sup>	37.07	85.4	EC	Re 61.13	$\gamma, e^-$
Au <sup>195</sup>	185 d	Pt <sup>194</sup> (d,n)Au <sup>195</sup>	32.8	—	EC	Pt 66.82	$\gamma, e^-$
Tl <sup>202</sup>	11.5 d	Hg <sup>201</sup> (d,n)Tl <sup>202</sup>	13.22	71.8	EC	Hg 70.82	$\gamma, e^-$
Tl <sup>204</sup>	2.7 y	Tl <sup>203</sup> (n, $\gamma$ )Tl <sup>204</sup>	29.50	96.0	EC, $\beta^-$	Hg 70.82	$\beta^-$
Bi <sup>207</sup>	50 y	Pb <sup>206</sup> (d,n)Bi <sup>207</sup>	23.6	81.0	EC	Pb 74.96	$\gamma, e^-$

state and decays to the normal state with internal conversion. Thus, two X-ray photons, characteristic of the same element ( $Z-1$ ), may be successively emitted. Cd<sup>109</sup> and Pd<sup>103</sup> are examples of these "double emitters".

**Photon flux density of nuclear X-ray sources**

The usefulness of a nuclear X-ray source generally will depend on the photon flux density of the radiation obtained, i.e. the number of X-ray photons emitted per sec and per cm<sup>2</sup>. Given a sample of

infinite thickness (see below) of a pure nuclide, the flux density is a characteristic of the material. For iron  $\text{Fe}^{55}$  e.g. it is  $25 \times 10^{10}$  photons/sec  $\text{cm}^2$ .

Some factors governing this characteristic may be indicated. The first factor is the specific activity, i.e. the number of disintegrations per second occurring in one gram of the pure nuclide. It is measured in curie/gram (c/g), one curie corresponding to  $3.7 \times 10^{10}$  disintegrations/sec. For isotopically pure  $\text{Fe}^{55}$ , the specific activity would be  $2.2 \times 10^3$  c/g. Other factors are: the branching ratio, i.e. the fraction of the total number of disintegrations due to any one of the X-ray producing mechanisms described; the internal conversion coefficient mentioned above, when internal conversion is the X-ray producing mechanism; the fluorescence yield, i.e. the fraction of the created K (or L, ...) vacancies resulting in the emission of K (or L, ...) photons; and the absorption of the X-ray producing element for the characteristic radiation emitted ("self-absorption" coefficient  $\mu_K$  or  $\mu_L$ , ...). Owing to the latter factor an increase in thickness of a nuclide layer does not provide a proportional increase in radiated flux density but a saturation is approached so that a source of a certain minimum thickness will have practically the same flux density as the infinite source.

The preparation of a nuclear X-ray source in usable form starts from a powdered, solid or dissolved specimen material, supplied by an atomic pile or a cyclotron<sup>6</sup>). This material, in general, contains the desired nuclide not in a pure form but mixed with other isotopes of the same element, which cannot be separated by normal laboratory means. The useful specific activity, i.e. the activity per gram of the element, is usually greatly lowered by these admixtures. Moreover, the other isotopes may give rise to contamination of the X-rays by undesired radiations. In order to minimize these adverse effects, the target material in the pile or cyclotron should contain the useful nuclide — from which the radionuclide in question is generated — in the highest possible concentration (target of increased abundance, cf. Table I). Obviously a high flux of the bombarding particles is also favourable for attaining high specific activities, since it will increase the theoretically attainable level of specific activity, at which equilibrium exists between the formation of new nuclei and the disintegration of nuclei already formed. The latter advantage is particularly important when the desired nuclide is contaminated by radioactive isotopes having a longer half-life: in that case, it may be desirable to keep the actual specific

activity well below the theoretically attainable level (by reducing the irradiation period of the target) in order not to get an X-ray source unduly contaminated by unwanted radiations.

The specific activities normally commercially available are given in Table II for a number of

Table II. Specific activity of normally available specimens of X-ray emitting nuclides, and X-ray photon flux density of infinite thickness samples.

Nuclide	Specific activity (millicurie per gram of element)	Photon flux density ( $10^7$ photons per sec. $\text{cm}^2$ )
$\text{Cr}^{51}$	400	2.85
$\text{Fe}^{55}$	1500	16.7
$\text{Ni}^{59}$	0.005	0.000087
$\text{Zn}^{65}$	5000	122
$\text{Ge}^{71}$	700	20.5
$\text{Se}^{75}$	100	3.25
$\text{Sn}^{113}$	1.6	0.283
$\text{In}^{114}$	720	129
$\text{Tm}^{170}$	10000	330
$\text{Tl}^{204}$	300	6.1

nuclides important as X-ray sources. Higher specific activities may be obtained in many cases at higher cost and longer delay in delivery. The last column of the table gives the calculated X-ray photon flux densities of infinitely thick samples of these specimens.

For an appreciation of these values consider the case of  $\text{Fe}^{55}$ , on which most of the nuclear X-ray work in this laboratory was done. The  $\text{Fe}^{55}$ -sources are usually used in a configuration shown in fig. 1.

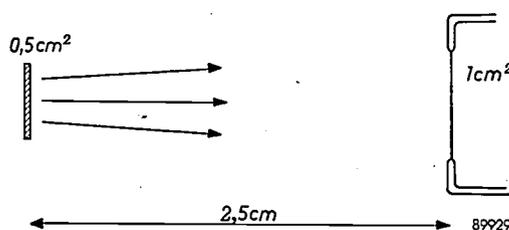


Fig. 1. Usual configuration of nuclear X-ray source (left) and counter tube under test (right).

The X-rays are emitted from a source of  $0.5 \text{ cm}^2$  area and enter the window of the detector (counter tube) to be checked or calibrated. The window area in most cases was  $1 \text{ cm}^2$ , the distance  $2.5 \text{ cm}$ . Under these conditions, the counter window encompasses about 1% of the radiation emitted by the source in all directions. The counting rate (at 100% counting efficiency of the detector) to be expected from a  $0.5 \text{ cm}^2$  source of infinite thickness of  $\text{Fe}^{55}$  having a flux density of  $16.7 \times 10^7$  photons/sec.  $\text{cm}^2$  would therefore

be  $8.4 \times 10^5$  counts/sec. With a "self-absorption" coefficient  $\mu \approx 700 \text{ cm}^{-1}$  of the source, i.e. the absorption coefficient of iron for the emitted  $\text{MnK}\alpha$  radiation ( $2.1 \text{ \AA}$ ,  $5.90 \text{ keV}$ ), it is calculated that even a source  $0.03 \text{ mm}$  ( $0.0012''$ ) thick would give 90% of the maximum rate. Several of the  $\text{Fe}^{55}$  sources prepared in this laboratory from a batch of  $\text{Fe}^{55}$  of the above-mentioned specific activity approached a rate of  $8 \times 10^4$  counts/sec, i.e. 10% of the maximum rate (the corresponding thickness of the layer is calculated to be about  $0.002 \text{ mm}$  or  $0.0001''$ ). This counting rate is amply sufficient for practical purposes, since larger counting rates would in any case have to be attenuated in order not to choke the counting systems at present available<sup>8)</sup>.

For the calculation of the absorption effect use is made of Dixon's calculations for large gamma-ray sources<sup>8)</sup>. The configuration of fig. 1 approximately meets the conditions that all detected photons traverse the source in a direction perpendicular to its surface and that the source has equal thickness in all parts of the beam cross-section. Under these conditions and with the assumptions of a homogeneous source, no scatter in the source and no absorption in the source holder window, the self-absorption in a source of thickness  $D$  is expressed by the equation<sup>9)</sup>:

$$\frac{I}{I_0} = \frac{1}{\mu D} (1 - e^{-\mu D}).$$

$I$  and  $I_0$  are the flux densities with and without self-absorption respectively.  $I_0$  evidently will be proportional to the thickness of the source:

$$I_0 = G \times D$$

and hence

$$I = \frac{G}{\mu} (1 - e^{-\mu D}).$$

For  $D \rightarrow \infty$ , the flux density approaches the value  $G/\mu$ , which in fact is the value listed in Table II. From the last equation the flux density of a source of any thickness can be calculated.

An obvious question is whether the flux density of the  $\text{Fe}^{55}$  source or other nuclear X-ray sources would render them possible substitutes for X-ray tubes in various applications, such as diffractometry and radiography. In diffractometry, however, the utilization of the X-rays emitted by the source is enormously poorer than in the configuration of fig. 1, so that much larger flux densities of the source are needed to get adequate counting rates in the diffraction lines. To illustrate this point, it should be mentioned that a normal diffraction X-ray tube with a loading of only  $0.1 \text{ mW}$  would emit approximately the same total number of K-photons/sec ( $\sim 10^7$ ) as the best practical  $0.5 \text{ cm}^2$   $\text{Fe}^{55}$  source described here<sup>10)</sup>. Since the actual loading of diffraction tubes may amount to  $1 \text{ kW}$

and their effective source size is e.g.  $0.015 \text{ cm}^2$ , the  $\text{Fe}^{55}$  source is evidently inferior by a factor of more than  $10^8$ . In radiography, the advantage of the X-ray tube as a source is even more pronounced since the continuous spectrum which the tube emits and which is useless in diffraction analysis, is fully effective photographically. In this case the  $\text{Fe}^{55}$  source would be equivalent to an X-ray tube with a specific loading of only appr.  $0.5 \mu\text{W}/\text{mm}^2$ . This should be compared to the actual loading applied in normal radiography, which may be  $4 \times 10^8$  times higher ( $200 \text{ W}/\text{mm}^2$ ) for a stationary anode and even  $10^{10}$  times higher ( $8000 \text{ W}/\text{mm}^2$ ) for a rotating anode (short exposures).

Obviously, then, the  $\text{Fe}^{55}$  source even at its theoretical maximum (i.e. with the radionuclide in a pure, undiluted state) cannot compete with the X-ray tube in flux density. Other radioactive nuclides may have higher theoretical flux density limits; however, because of the fundamental relationship between specific activity  $S$  and half-life  $T$  of a radioactive element,

$$S = \frac{4.16 \times 10^{23}}{A \times T}$$

( $T$  being measured in seconds and  $A$  being the atomic weight), flux densities comparable to those of X-ray tubes will be attainable only with nuclides far too short-lived for practical purposes.

#### Preparation of the sources<sup>11)</sup>

A number of factors led us to select  $\text{Fe}^{55}$  as the main nuclide for our experimental X-ray sources.  $\text{Fe}^{55}$  was found suitable for the envisaged applications not only because of the wavelength of the emitted  $\text{MnK}\alpha$  radiation but also by virtue of its 3 year half-life and its decay by pure electron capture unaccompanied by processes generating other radiations. It has furthermore been shown in the above that the specific activity of the available  $\text{Fe}^{55}$  specimens permits the fabrication of sources of sufficient X-ray intensity for work on counter tubes.

The radioactive  $\text{Fe}^{55}$  is obtained from the suppliers in the form of an  $\text{FeCl}_3$  solution, from which  $\text{Fe}^{55}$  may be laid down in some solid, homogeneous form by means of a number of chemical and physical techniques. Three methods were successfully used. (Similar techniques are applicable to most other X-ray emitting nuclides.)

1) The filter paper method. Several drops of the active solution are transferred to a  $1 \text{ cm}^2$  piece of filter paper which has been positioned in a previously prepared holder. The assembly is then sealed with  $12 \mu$  polyester film, which is 98% transparent to the  $\text{MnK}\alpha$  radiation and which prevents the  $\text{FeCl}_3$  solution giving off  $\text{HCl}$ -vapor. This simple procedure produces a stable source, which, however, is rather weak owing to the diluted state of the radioactive nuclide and the absorption of X-rays by Cl atoms

<sup>8)</sup> W. R. Dixon, *Nucleonics* 8, No. 4, April 1951, pp. 68-72.

<sup>9)</sup> Cf. e.g. D. Taylor, *The measurement of radioisotopes*, Methuen and Co., London, 1951.

<sup>10)</sup> This calculation and the following ones are due to W. J. Oosterkamp of the Philips Laboratory at Eindhoven. Cf. also: W. J. Oosterkamp, *The radiography of the human body with radioactive isotopes*, *Brit. J. Radiol.* 26, 111, 1953.

<sup>11)</sup> The authors gratefully acknowledge the assistance of W. B. Brown of this laboratory in this phase of the work.

(mass absorption coefficient for 2 Å appr. 300) and by the filter paper.

2) The ammonia method. From a drop of the active  $\text{FeCl}_3$  solution placed on a small glass or metal plate,  $\text{Fe}(\text{OH})_3$  is precipitated by means of an ammonia atmosphere. Heating the plate then converts the iron hydroxide into the non-volatile iron oxide, which strongly adheres to the base material. No window is necessary and the absorption is substantially lower than in the previous case (the mass absorption coefficient of the oxygen atoms for 2 Å is only about 30). Even higher intensities can be obtained when the oxide is subsequently reduced in an  $\text{H}_2$ -atmosphere and, for protection, the metallic Fe fused into a platinum backing by heating. This method was developed at Oak Ridge National Laboratory<sup>12)</sup>.

3) Electroplating. This method, whereby pure  $\text{Fe}^{55}$  can be deposited in successive thin layers, results in maximum concentration of the radioactive material and in the highest photon flux densities attained. Conventional procedures for electroplating iron proved not amenable to the micro-techniques involved in the preparation of radioactive sources. A procedure was worked out after a suggestion by Platzer and Lewis<sup>13)</sup>, the main problem being to avoid etching away of the deposited iron by acid formed in the electrolysis of the  $\text{FeCl}_3$ -solution. Fig. 2 shows the equipment used. A small amount of the active solution, to which no "carrier" substance is added, is pipetted onto a filter paper placed on a graphite anode. A copper plate forming the cathode, on which the Fe is to be plated, is held on a movable fixture and is lowered into position on the filter paper, while a potential of 3 V is applied to the electrodes. After the deposition process, the whole structure is flooded with a spray of water to wash away the formed acid<sup>14)</sup>. The process can be repeated with a fresh  $\text{FeCl}_3$  portion, and in this way as many as ten successive layers have been plated onto one backing piece, building up a total thickness which produces 10% of the infinite-thickness flux density (see above).

Similar techniques have been evolved to deposit  $\text{Zn}^{65}$ . This radioactive nuclide is particularly

<sup>12)</sup> R. E. McHenry (Oak Ridge National Laboratory), unpublished report, Dec. 29, 1954.

<sup>13)</sup> G. E. Platzer and C. R. Lewis (Chrysler Corp., Detroit, Mich.), private communication.

<sup>14)</sup> Of course, the radioactively contaminated washing water must be collected and safely disposed of. We do not dwell on these and similar protective measures that have to be taken in order to safeguard the operators and surroundings in all procedures for preparing nuclear X-ray sources. It may, however, be mentioned that such measures are relatively few and easy in the electroplating method, since intermediate manipulation and processing of the material is not required.

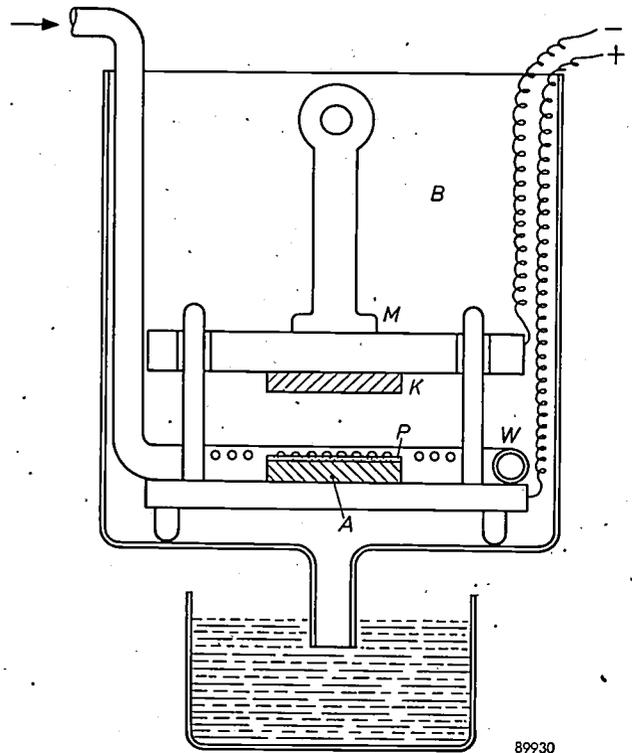


Fig. 2. Equipment for electroplating  $\text{Fe}^{55}$ . *K* copper plate on which  $\text{Fe}^{55}$  is to be deposited, held by the movable fixture *M* in glass beaker *B*. *P* filter paper placed on graphite anode *A* and saturated with radioactive  $\text{FeCl}_3$  solution. A 3V potential is applied between *K* and *A* and the cathode *K* is then lowered onto the filter paper. After the plating process, wash water is sprayed on the electrodes from a perforated ring *W* before the electrodes are separated.

important since it emits  $\text{CuK}\alpha$  rays ( $\lambda = 1.54 \text{ \AA}$ , energy 8.05 keV). From the tables it is seen that the half-life of this source is 250 days and that the infinite-thickness flux density of the best specimens at present obtainable is  $122 \times 10^7$  photons/sec  $\text{cm}^2$ .

The design of the holder for the nuclear X-ray source depends on the intended application, since it should provide means for proper positioning in the

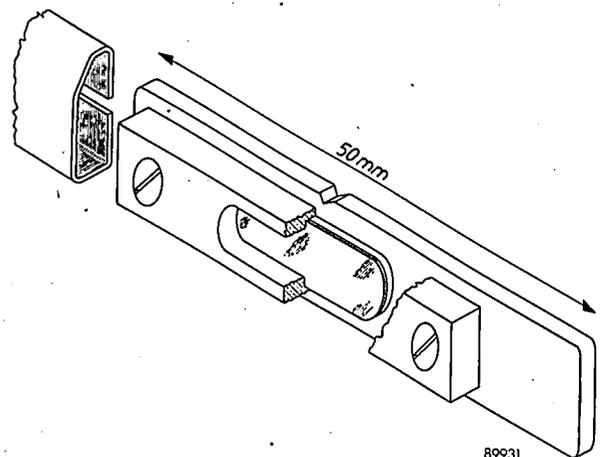


Fig. 3. Holder for nuclear X-ray source designed for location in the position of the scatter slit of the "Norelco" X-ray goniometer. To the left, part of a protective cover for the source when not in use.

equipment in which the source is to be used. General requirements are easy portability, adequate shielding during use and storage and a design such as to prevent the dispersal of the active material within or outside the holder. For many laboratory applications a long-handled holder with a remote control shutter mechanism has proved very useful. For users of the "Norelco" equipment for X-ray diffractometry and X-ray spectrochemical analysis<sup>15)</sup> the holder shown in *fig. 3* will be of foremost interest. In this holder, the source is mounted on a brass strip that can be located in the position of the scatter slit of the "Norelco" goniometer, i.e. immediately in front of and at an accurately reproducible distance from the X-ray detector (Geiger counter, proportional counter, scintillation counter), as shown in *fig. 4*.

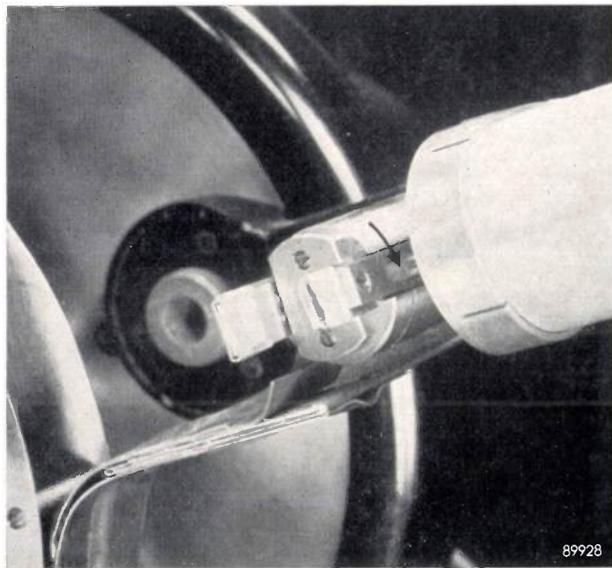


Fig. 4. Detector carrying arm of "Norelco" X-ray goniometer showing the detector to the right and the source of *fig. 3* in position in the scatter-slit holder (arrow).

### Applications

Sources as shown in *fig. 3* have been extensively used in our laboratories for checking the stability of the complete detecting system of the above-mentioned "Norelco" equipment. The nuclear source by virtue of its ideal stability is a very effective auxiliary for testing and singling out causes of instability in such equipment. Owing to the 3 year half-life of  $\text{Fe}^{55}$ , its radiation output shows a fractional decrease of only 0.07% per day, which moreover can be easily corrected for, since the fractional disintegration rate is constant and

accurately known. Output fluctuations due to the statistical nature of the photon emission processes are low enough, with the high intensity source described, to preclude difficulties in this application.

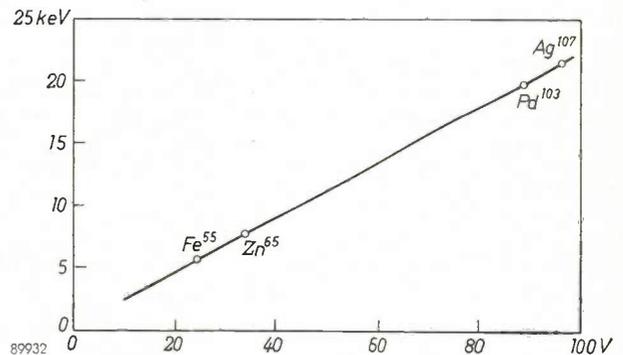


Fig. 5. Calibration of proportional counter. The photon energy (in keV) is plotted against the mean pulse amplitude (in volts), using four points corresponding to the radiation of the nuclides  $\text{Fe}^{55}$ ,  $\text{Zn}^{65}$ ,  $\text{Pd}^{103}$  and  $\text{Ag}^{107}$ .

The monochromatic emissions of nuclear X-ray sources, such as  $\text{Fe}^{55}$  and  $\text{Zn}^{65}$ , have made them very useful for checking the proportionality between pulse amplitude and photon energy in the proportional and scintillation counter. At the same time a calibration of the counter enabling the identification of quanta from measured pulse amplitudes is obtained (*fig. 5*), and recordings can be made of the pulse amplitude distribution curve (*fig. 6a, b*) on which depends the discrimination ability of the counter<sup>3)</sup>. The degree of linearity of the relation between counting rate and irradiation intensity, and the "plateau" characteristic of Geiger and proportional counters are other characteristics of counters which can be easily tested with the aid of a nuclear X-ray source.

All these tests, using nuclear sources, are performed during the actual manufacture of "Norelco" radiation detectors<sup>3)</sup>. Thus proportional counters are checked before being sealed off, and the total gas pressure is adjusted so as to give a specified pulse amplitude at a given operating voltage on irradiation with a  $\text{Fe}^{55}$  source. The components of scintillation detectors, i.e. the crystal and the photomultiplier tube, are tested separately on the work bench. Finally, life-tests of groups of detectors at high counting rates are effected by means of a number of  $\text{Fe}^{55}$  sources, thus avoiding the tying-up of costly X-ray tube equipment for long periods.

The applications of nuclear X-ray sources so far mentioned are concerned with the testing of detectors. Applications in other fields, especially those in which monochromatic radiations are required, have been considered or tried out already. X-rays

<sup>15)</sup> Philips tech. Rev. 16, 123-133, 1954/55; 17, 269-286, 1955/56.

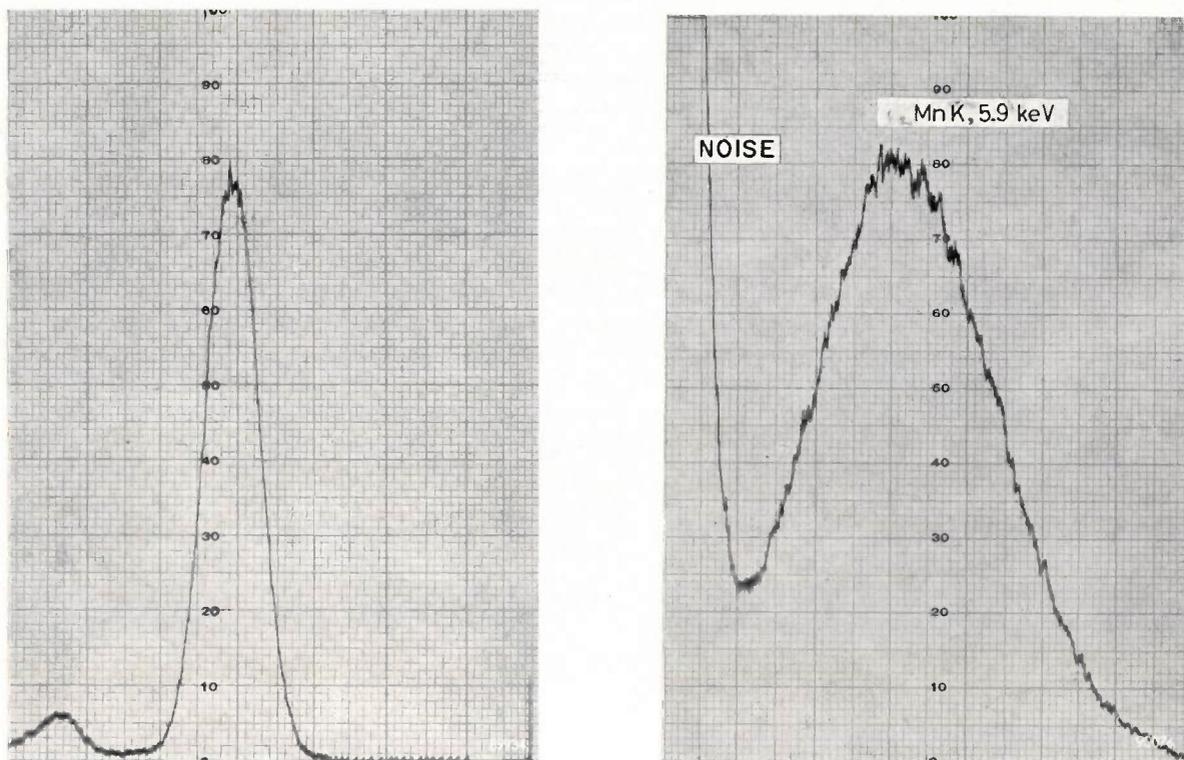


Fig. 6. a) Pulse amplitude distribution of a proportional counter irradiated with Mn K $\alpha$  rays from an Fe<sup>55</sup> source.

b) The same for a scintillation counter.

The curves were obtained with a pulse height analyzer channel width ( $\Delta V$ ) of  $\frac{1}{10}$  of the half-width of the peak (cf. <sup>3</sup>): the number of pulses transmitted via the different channels were directly recorded over the corresponding channel position (pulse amplitude).

from radioactive nuclides have been used as wavelength standards<sup>16</sup>); absorption measurements of the emission of two adjacent nuclides have been proposed for quantitative analysis of an element having an absorption edge in the same wavelength region<sup>17</sup>); absorption analysis of sulfur in hydrocarbons has been accomplished with MnK $\alpha$  radiation from a Fe<sup>55</sup> source<sup>18</sup>). Application of radionuclides such as Tm<sup>170</sup> (thulium) emitting photons of energy 50 and 80 keV have been described, e.g. for inspection in industry and for emergency medical radiography<sup>2</sup>)<sup>19</sup>). Thickness gauging is another possible application in industry. Further work in the development of nuclear X-ray sources is warranted in order to broaden the field of application of X-rays by simple,

inexpensive techniques for cases where the very high photon flux densities of X-ray tubes are not required.

**Summary.** A number of artificial radioactive nuclides emit photons of energy less than 100 keV, i.e. of wavelength longer than 0.15 Å (relatively soft X-rays). Such nuclear X-ray emitters owing to their high stability, monochromatic emission and low cost provide excellent means for testing X-ray counter tubes (especially proportional and scintillation counters), both during manufacture and in operation. After a discussion of the underlying mechanisms of the X-ray emissions, a survey of X-ray emitting nuclides with their main characteristics is given, and some techniques for preparing practically usable sources from these nuclides are described. The theoretically attainable photon flux density of Fe<sup>55</sup> would provide a counting rate of nearly 10<sup>6</sup> counts/sec in the envisaged application; the best practical Fe<sup>55</sup>-sources prepared in the Philips Laboratories attain about 10% of this value, which is amply sufficient. Fe<sup>55</sup>, Zn<sup>65</sup> and other X-ray sources made in these laboratories have for some time been used for testing the stability of the detection system of "Norelco" X-ray goniometers, for checking and calibrating the photon energy proportionality, the pulse amplitude distribution, the counting-rate linearity and the plateau characteristics of "Norelco" proportional and scintillation counters, for life-testing these counters etc.

Nuclear X-ray sources cannot compete on equal terms with X-ray tubes, whose useful photon flux densities may be 10<sup>8</sup> or in some cases 10<sup>10</sup> times higher; nevertheless, they may be important for a broad and diverse field of applications in which monochromatic emissions are desired and high flux densities are not necessary.

<sup>16</sup>) Ph. Snelgrove, J. El-Hussaini and J. W. M. DuMond, Phys. Rev. **95**, 1203-1204, 1954.

<sup>17</sup>) Suggested by H. A. Liebhafsky, in paper presented at Pittsburgh Conf. on Anal. Chem. and Appl. Spectroscopy, March 1955.

<sup>18</sup>) H. K. Hughes and J. W. Wilczewski, Anal. Chem. **26**, 1889-1893 1954.

<sup>19</sup>) S. Untermyer et al., Nucleonics **12**, No. 5, May 1954, pp. 35-37; R. Halmshaw, Brit. J. appl. Phys. **6**, 8-10, 1955.

**ELECTRON MICROSCOPE PHOTOGRAPH OF FERROXDURE**

Surface of a block of ferroxdure containing a slight excess of BaO, sintered at 1200 °C, sawn through and then reheated to a somewhat higher temperature. Photographs such as this can provide information on diffusion and grain growth in the process of recrystallization.

The photograph was made with a carbon replica (shadowed with palladium), i.e. a carbon skin evaporated on to a polystyrene impression prepared from the unpolished surface of the ferroxdure. The photograph was taken with an electron microscope EM 75 kV with an anastigmat lens<sup>1)</sup>. Magnification 23 000×.

<sup>1)</sup> Philips tech. Rev. 17, 47-59, 1955/56. The anastigmat developed for this microscope, which improves the resolving power to about 25 Å, will be described later in this Review.

## A TRANSISTOR PRE-AMPLIFIER FOR THE MAGNETODYNAMIC PICK-UP

by C. HUBER and J. RODRIGUES de MIRANDA.

621.375.4:681.84.081.48

*The new, magnetodynamic pick-up<sup>1)</sup> requires a pre-amplifier if it is to be used in combination with the low-frequency section of a radio set. By employing a transistor the authors have succeeded in housing the pre-amplifier, complete with its supply unit, in a box about  $3\frac{1}{2}'' \times 3'' \times 2''$ .*

A new pick-up (type AG 3020/21) working on the magnetodynamic principle, has recently been marketed and has been the subject of two articles in this journal<sup>1)</sup>. It may be recalled that this pick-up employs a ferroxdure magnet in the form of a thin rod, magnetized at right-angles to its axis and mounted so as to be able to rotate about its axis, between the ends of a magnetically conducting yoke carrying two coils. The vibrations of the needle are converted, via a lever (the needle arm), into a turning movement of the magnet about its axis. This movement sets up an alternating flux in the yoke, which induces an alternating voltage in the coils.

The magnetodynamic pick-up is of such a quality that it can compete with electrodynamic pick-ups, which have hitherto generally been used for professional purposes. On the other hand its price is not so high as to prohibit its use in non-professional record players. In the latter case the pick-up will often be used in conjunction with a radio set, as is commonly the case with piezo-electric pick-ups. There are, however, two reasons why direct connection of the magnetodynamic pickup to an ordinary radio would not give satisfactory results. Firstly, it is less sensitive, so that an extra amplification stage is necessary. Secondly, the voltage delivered is proportional not to the amplitude, as in the case of the piezo-electric pick-up, but to the velocity of the needle point. This difference necessitates an electrical correction, which can be applied in the extra amplification stage.

Some correction is necessary even with the piezo-electric pick-up. Often this is nothing more than a very rough correction by means of mechanical resonances in the pick-up system. The better pick-ups, such as the type AG 3015 piezo-electric pick-up and the magnetodynamic pick-up, demand a more accurate correction, which it is practically impossible to obtain otherwise than electrically.

It is desirable that the extra amplification stage (pre-amplifier incorporating correction devices) be

easily interposed between the magnetodynamic pick-up and a normal radio; it should therefore preferably be small and unobtrusive. Still more important, it should include its own supply unit, in order to avoid the complication of drawing the supply from the radio.

Adequate amplification could easily be attained with one valve, e.g. a double triode, but transistors are much more convenient for the present purpose: they are smaller than valves, have no filament, and require less D.C. power. This power can then easily be supplied by the mains via a germanium diode, which compared with a rectifying valve has the same advantages which transistors have over amplifying valves, viz. it is smaller in size and has no filament.

One transistor is found to be adequate. The complete circuit — the signal current and direct current parts of which will be examined separately later in the article — is shown in *fig. 1*. It will be observed that the transistor *T* is used in the circuit with a common emitter, in which the base forms the input electrode and the collector the output electrode. The pre-amplifier (type AG 9005) in its box, is shown in *fig. 2*.

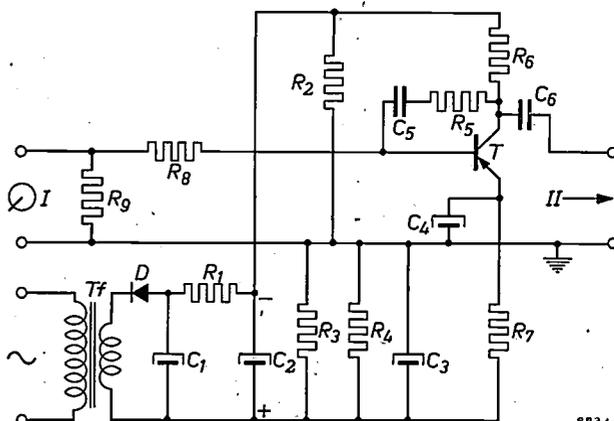


Fig. 1. Circuit of the pre-amplifier type AG 9005. The input terminals *I* are connected to a magnetodynamic pick-up AG 3020 or AG 3021, the output terminals *II* to the pick-up terminals of a radio set. *T* junction transistor OC 73. *T<sub>f</sub>* supply transformer. *D* germanium diode OA 81. *C*<sub>1</sub>-*R*<sub>1</sub>-*C*<sub>2</sub> smoothing circuit. *C*<sub>3</sub>, *C*<sub>4</sub> decoupling capacitors. *C*<sub>6</sub> coupling capacitor. For the significance of the other letters, see *fig. 4* and *fig. 6*.

<sup>1)</sup> N. Wittenberg, A magnetodynamic gramophone pick-up: I. Construction, II. Frequency characteristics, Philips tech. Rev. 18, 101-109, and 173-178, 1956/57.

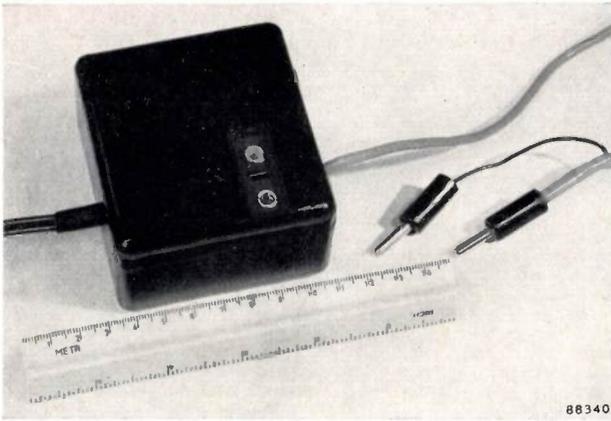


Fig. 2. External appearance of the pre-amplifier (with power unit) type AG 9005. Approximate dimensions: base  $3\frac{1}{2}'' \times 3''$ , height  $2''$ .

**Frequency characteristic**

For most gramophone records the recording curve is of the type shown by *B* in fig. 3. This "New Orthophonic" curve, as it is called, gives the peak stylus velocity  $\hat{v}$  with which the groove is cut in the

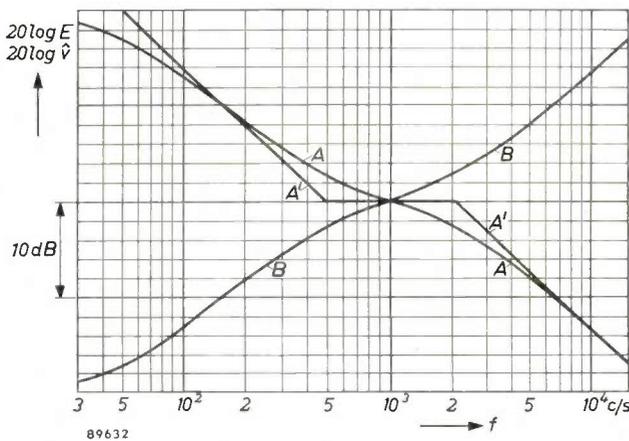


Fig. 3. *A* American "Standard Response" curve. *B* "New Orthophonic" recording characteristic. *A'* schematic approximation to curve *A*.

lacquer disc during recording<sup>2)</sup> as a function of the frequency  $f$  on a logarithmic scale. *B* is the mirror image of the "standard response" curve *A* with respect to the level at 1 kc/s; the "standard response" curve gives the output voltage  $E$  as a function of  $f$  of a certain reproducing channel proposed as a standard for gramophone reproduction, the needle point of the pick-up having a constant peak velocity. With the magnetodynamic pick-up, the induced e.m.f. is proportional to the velocity of the needle

point and thus (if the pick-up is assumed to be perfect) to the velocity of the cutting stylus. The amplifier used with the magnetodynamic pick-up must, therefore, have a frequency characteristic which approximates as closely as possible to the standard response curve *A*. In the case in question, it is the characteristic of the pre-amplifier which must conform to the curve *A*, since the remaining part of the reproduction channel (i.e. the low-frequency section of the radio) is designed for signals which are independent of the frequency. Frequencies above 1 kc/s must therefore be amplified by the pre-amplifier to a lesser extent than the frequencies below 1 kc/s.

A good approximation to the desired characteristic has been obtained by means of two provisions: 1) a negative feedback link  $R_5$ - $C_5$  between the collector and the base, and 2) a resistor  $R_8$  in series with the pick-up (see fig. 4, which shows only the signal current section of fig. 1; the pick-up is now represented as a generator of e.m.f.  $E_i$ , self-induction  $L_i$  and internal resistance  $R_i$ ).

To illustrate the effect of these measures, consider now in fig. 3 the schematic curve *A'* in place of the standard response *A*. This exhibits an amplification which decreases inversely with the frequency up to 0.5 kc/s and again above 2.1 kc/s, and is constant between these two frequencies.

At frequencies  $f = \omega/2\pi$  such that  $1/\omega C_5 \ll R_5$ , the link  $R_5$ - $C_5$  produces a negative feedback which is (virtually) independent of  $f$ , so that in this range the amplification (apart from other effects) is independent of  $f$ . At low frequencies, when  $1/\omega C_5 \gg R_5$ , on the other hand, the negative feedback increases in direct proportion to  $f$ , and consequently the amplification decreases approximately inversely with  $f$ , just as required by the left-hand side of the schematic response curve *A'*. The transition from the one frequency range to the other is characterized

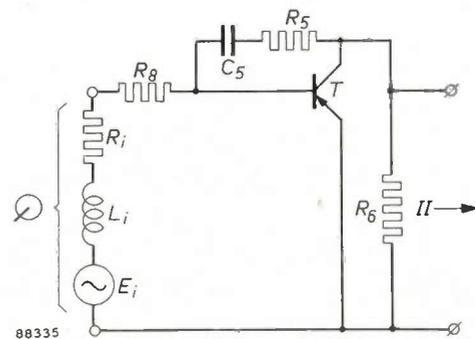


Fig. 4. The signal-carrying sections of the circuit of fig. 1.  $E_i$ - $L_i$ - $R_i$  equivalent circuit of the pick-up.  $R_8$  series resistor which causes relatively smaller amplification at high frequencies.  $R_5$ - $C_5$  negative feedback, providing relatively greater amplification at low frequencies. *T* transistor OC 73.  $R_6$  collector resistor. *II* output terminals.

<sup>2)</sup> J. L. Ooms, The recording and production of gramophone records, Philips tech. Rev. 17, 101-109, 1955/56, particularly p. 104.

by the frequency  $f_1 = 1/(2\pi R_5 C_5)$ . The values of  $R_5$  and  $C_5$  are chosen in accordance with the response curve so that  $f_1$  is about 0.5 kc/s.

The combination of inductance and resistance in the circuit comprising  $L_i$ ,  $R_i$ ,  $R_g$  and the base-emitter resistance  $R_{be}$  has just the opposite effect. The alternating current in the base is (virtually) independent of  $f$  for frequencies  $f \ll f_2 = (R_i + R_g + R_{be})/2\pi L_i$ , and is (almost) inversely proportional to  $f$  for  $f \gg f_2$ . If we now plot the logarithm of the base current against  $\log f$ , we find therefore a horizontal line for  $f \ll f_2$ , and a falling straight line above this frequency, corresponding to the right-hand part of the response curve. The transition frequency lies at 2.1 kc/s. The resistance  $R_g$  is so chosen that  $f_2$  has this value.

The result of both the above measures combined is shown in fig. 5, where the output voltage  $E_o$  of the pre-amplifier is plotted as a function of  $f$ , the

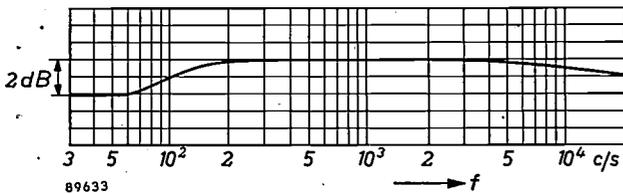


Fig. 5. Output voltage  $E_o$  on a logarithmic scale as a function of the frequency  $f$ ,  $E_i$  varying with  $f$  in accordance with the characteristic  $B$  in fig. 2. External load 150 k $\Omega$ .

input voltage  $E_i$  varying with  $f$  in accordance with the "New Orthophonic" recording characteristic ( $B$  in fig. 3). The variation of  $E_o$  is less than 2 dB in the frequency range from 40 to 20 000 c/s.

Data concerning sensitivity, amplification and distortion of the pre-amplifier in combination with the magnetodynamic pick-up are given in the table below. Certain quantities quoted in the table may be usefully defined here.

Table. Sensitivity, amplification and distortion data on the pick-up AG 3020/21 with the pre-amplifier AG 9005. External load: resistance  $\geq 120$  k $\Omega$  in parallel with a capacitance of  $\leq 10^{-9}$  F.

Frequency	60 c/s	1000 c/s	10 000 c/s	
Sensitivity . . . . .	400	71.5	15.5	mV/(cm/s)
Voltage amplification . . .	41	27	25	dB
Current amplification . . .	21.4	7.8	5.2	dB
Insertion gain . . . . .	41	25	13.6	dB
Transfer impedance . . . .	1170	230	180	mV/ $\mu$ A
Distortion (for maximum $E_o$ and an ambient temperature of 20 $^{\circ}$ C) . . .	$\leq 1.5$	$\leq 0.4$	$\leq 2$	%

*Sensitivity*: output voltage  $E_o$  (r.m.s. value) in mV for a peak needle velocity of 1 cm/sec.

*Voltage amplification*:  $20 \log E_o/E_i$  dB.

*Current amplification*:  $20 \log I_o/I_i$  dB ( $I_i$  = input current,  $I_o$  = output current).

*Insertion gain*:  $20 \log E_o/E_i$  dB, where  $E_i$  is the terminal voltage of the pick-up when the latter is loaded with its normal load resistance (68 k $\Omega$ ).

*Transfer impedance*:  $E_o/I_i$ , in mV/ $\mu$ A.

**Stabilization of the working point**

At maximum modulation the output voltage may show peak values of 3.5 V. To avoid distortion it is necessary to ensure that the D.C. voltage  $|V_{ce}|$  between collector and emitter never falls below 4V.  $|V_{ce}|$  is given the rather high nominal value of 10 V, but for a number of reasons the danger that  $|V_{ce}|$  might become less than 4 V is a real one: not only must mains voltage fluctuations be taken into account, but also the quite appreciable effect of the temperature, the spread in the properties of transistors and the spread in resistance values.

A new type of junction transistor<sup>3)</sup> has been selected, OC 73, which has been found during manufacture to have an appreciably smaller spread in properties than the older types. Resistors with a tolerance of  $\pm 2\%$  have been specified where the accurate value of the resistance is critical.

The effect of the temperature still remains. In general, temperature has quite an appreciable effect on transistors, and the pick-up preamplifier is required to work well for all ambient temperatures from 10 to 45  $^{\circ}$ C (this higher temperature limit is stipulated since it is anticipated that the magnetodynamic pick-up will find a ready market in the tropics as it is better able to withstand tropical conditions than are other systems).

As a first step towards minimizing the effect of temperature a resistor  $R_7$  is connected in series with the common electrode — in this case the emitter (see fig. 6, which shows the direct current part of the circuit in fig. 1). Just as a cathode resistor has a stabilizing action on the anode current of valves, so here  $R_7$  (shunted by capacitors  $C_3$  and  $C_4$  in series, see fig. 1) has a stabilizing effect on the collector D.C. (D.C. negative feedback), and consequently on the voltage  $V_{ce}$ .

This single precaution is not enough however. In order to obtain high amplification, a high value has been chosen for the collector resistance  $R_6$  (0.1 M $\Omega$ ). Hence a slight increase in the collector

<sup>3)</sup> For the simple theory of the junction transistor, see F. H. Stieltjes and L. J. Tummers, Philips tech. Rev. 17, 233-246, 1955/56.

current (as a result of an increase in temperature) may be sufficient to cause  $|V_{ce}|$  to fall below the admissible value. For this reason, as a second stabilizing device, a thermistor (NTC resistor)  $R_3$  with a large negative temperature coefficient <sup>4)</sup> has been

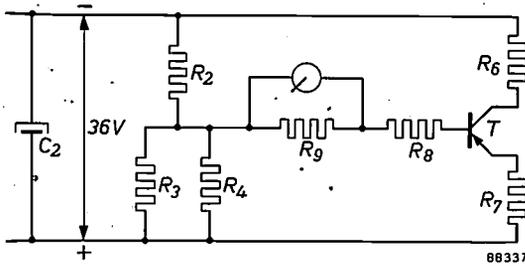


Fig. 6. The D.C. sections of the circuit of fig. 1.  $C_2$  smoothing capacitor.  $R_2 + (R_3 // R_4)$  voltage divider ( $R_3$  thermistor,  $R_4$  ordinary resistor).  $R_6$  collector resistor.  $T$  transistor OC 73.  $R_7$  resistor for D.C. negative feedback.  $R_8$  series resistor (see fig. 4).  $R_9$  resistor which limits the voltage across the transistor when the pick-up is disconnected.

incorporated in the potential divider  $R_2 + (R_3 // R_4)$  which determines the potential of the base. As a result the potential of the base is dependent on the ambient temperature in such a way that the effect of the temperature on the transistor is adequately compensated.

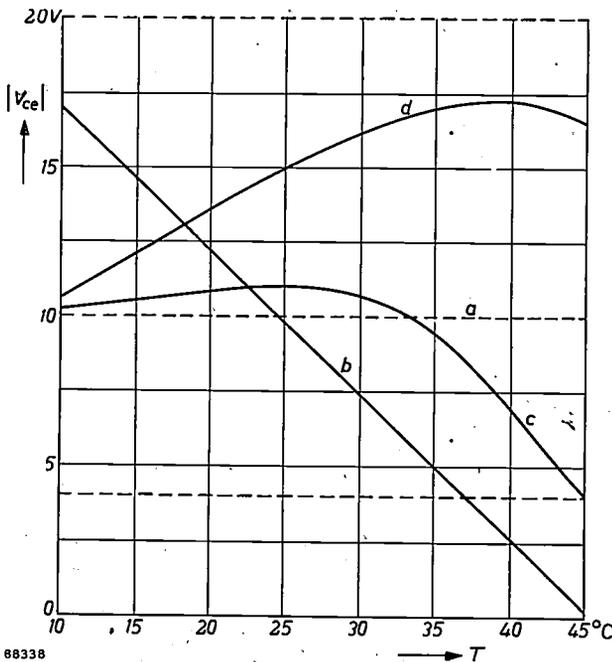


Fig. 7. Line *a*: constant value of 10 V at which it would be desirable to keep  $|V_{ce}|$ , irrespective of the ambient temperature  $T$ . Line *b*: variation of  $|V_{ce}|$  in absence of thermistor ( $R_3$  in fig. 6). Lines *c* and *d*: extreme values assumed by  $|V_{ce}|$  (all tolerances having been taken into consideration), when an ordinary resistor ( $R_4$ ) is used in parallel with a thermistor ( $R_3$ ).

This compensation can be explained as follows. In fig. 7, *a* indicates the value at which it would be desirable to keep  $|V_{ce}|$  constant (10 V). If there were no thermistor,  $|V_{ce}|$  would depend upon the ambient temperature  $T$  as shown by line *b*. If  $R_3 // R_4$  is replaced by a variable resistor and a graph is plotted of how resistance must vary to keep  $|V_{ce}| = 10$  V as  $T$  changes from 10 to 45 °C, the curve *R* in fig. 8 is obtained. The variation of a thermistor with  $T$  differs considerably from curve *R*, as shown by curve *NTC*. By connecting a thermistor in parallel with an ordinary resistor, a temperature variation like  $R_3 // R_4$  in fig. 8 can be obtained which can be made to approximate sufficiently closely to the desired variation (*R*). In the case in point this is achieved by connecting a thermistor of 47 kΩ (at 25 °C) in parallel with an ordinary resistor of 20 kΩ. The extreme values of  $|V_{ce}|$ , found with this combination, all tolerances having been taken into consideration, are given by the curves *c* and *d* of fig. 7. It may be seen that in the temperature range from 10 to 45 °C,  $|V_{ce}|$  never falls below 4. V and never exceeds 20 V (the maximum value which is permissible for the transistor).

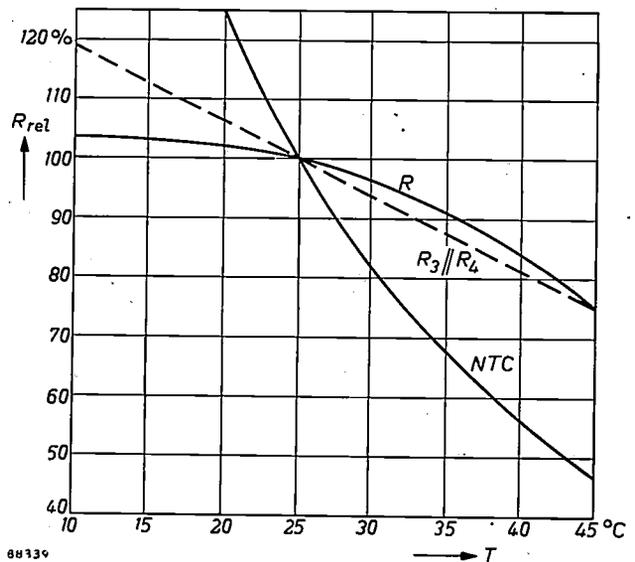


Fig. 8. *R*: variation of the resistance of  $R_3 // R_4$  in fig. 6 required if  $|V_{ce}|$  is to remain equal to 10 V for  $T = 10$  to 45 °C. *NTC*: variation of the resistance value of a thermistor.  $R_3 // R_4$  approximation to the curve *R*, obtained by connecting a thermistor and an ordinary resistor in parallel. (All resistance values are on a relative scale:  $R = 100$  at 25 °C.)

Resistor  $R_9$  (fig. 1 and fig. 6) ensures that the base circuit is not interrupted when the pick-up is being changed or disconnected and prevents too high a voltage from being applied to the transistor. The resistance of  $R_9$  is so high (0.47 MΩ) that it has no effect when the pick-up is plugged in.

It may be noted that the D.C. base current (maximum about 50 μA) flows through the pick-up, but is quite harmless.

**Power unit**

The high values of  $|V_{ce}|$  (nominal 10 V) and of the collector resistor  $R_6$  necessitate a relatively high supply voltage, namely 36 V. This voltage

<sup>4)</sup> E. J. W. Verwey, P. W. Haayman and F. C. Romeyn, Philips tech. Rev. 9, 239-248, 1947/48.

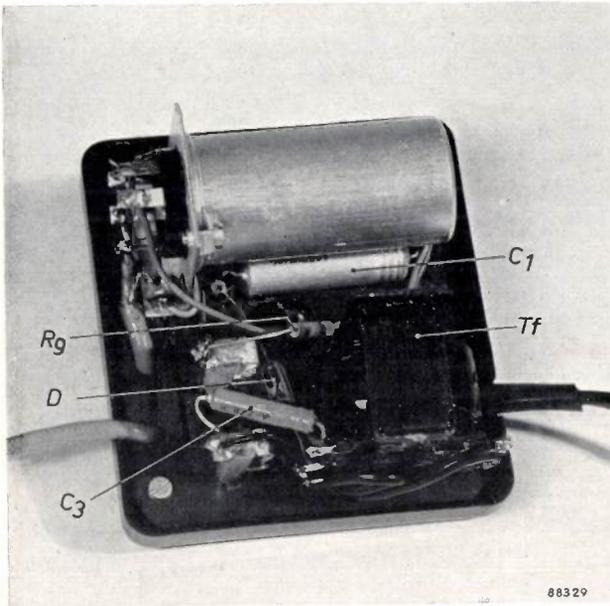


Fig. 9. The pre-amplifier AG 9005 with cover removed. The cylindrical screening can houses the pre-amplifier proper (see fig. 10). Significance of the letters as in figs. 1, 4 and 6

can be obtained by rectification with a germanium diode<sup>5)</sup> OA 81, the maximum peak inverse voltage of which is 115 V. With the current required here (about 0.4 mA) this diode is only carrying a fraction of its maximum permissible load.

The filter  $C_1$ - $R_1$ - $C_2$  (fig. 1) smooths the rectified voltage to the extent that the extra hum produced in the loudspeaker when the pre-amplifier is switched in is barely audible. This means that the hum component in the output voltage of the pre-amplifier is less than 1.5 mV, i.e. about 65 dB below the maximum output voltage.

There is one point to be mentioned regarding the transformer. In conformity with safety requirements the supply transformers in radio sets, amplifiers, etc. are provided with a temperature-sensitive safety device which automatically cuts off the mains supply to the transformer if the latter threatens to become too hot. In the pre-amplifier under discussion, the transformer is so small that it would be difficult to fit such a safety device. Fortunately such a device is found to be unnecessary since the transformer is short-circuit proof. In normal operation the power dissipated mainly comprises iron losses. If the secondary coil is short-circuited the copper losses of course increase, but owing to the relatively

large stray inductance and resistance of the coil the increase is only small. Moreover, owing to the increased voltage drop in the primary coil, the induced e.m.f. becomes smaller, and hence the iron losses decrease. Thus in the event of a short circuit the total loss changes hardly at all, so that the transformer does not become too hot.

### Mechanical construction

The pre-amplifier complete with power unit is mounted in a rectangular "Philite" box (fig. 2 and fig. 9), of size  $86 \times 78 \times 49$  mm (about  $3\frac{1}{2}'' \times 3'' \times 2''$ ). The pre-amplifier proper is constructed as a self-contained unit, all its components being mounted on an insulating plate which is housed in a cylindrical screening can (fig. 10). The reason for this is that if future designs of radio sets and amplifiers should have the pre-amplifier power supply (36 V, 0.4 mA) already built-in, it will only be necessary to add the pre-amplifier proper as in fig. 10.

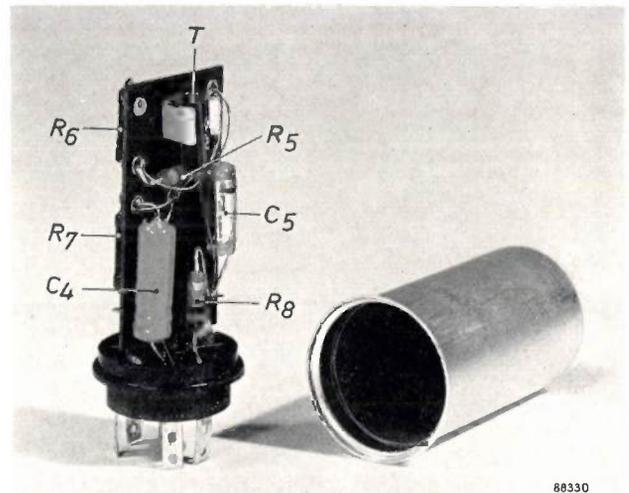


Fig. 10. The pre-amplifier proper with screening can (diameter 1 inch, length 2 inches). Significance of the letters as in figs. 1, 4 and 6.

**Summary.** Description of the pre-amplifier AG 9005, designed for coupling the magnetodynamic pick-up AG 3020/21 to a normal radio. Its purpose is to amplify the signal voltage to an adequate level (max 2.5 V r.m.s.) and at the same time to provide a frequency characteristic conforming to the "New Orthophonic" cutting characteristic of gramophone records. The necessary amplification is obtained with one OC 73 transistor connected with common emitter. As one of the measures taken to stabilize the working point at differing ambient temperatures (10 to 45 °C), a thermistor is incorporated. The power unit contains a transformer, a germanium diode OA 81 and a smoothing filter. The pre-amplifier proper, together with supply unit, is mounted in a box about  $3\frac{1}{2}'' \times 3'' \times 2''$ .

<sup>5)</sup> Philips tech. Rev. 16, 225-232, 1954/55.

## A NEW KIND OF TUNING INDICATOR TUBE

621.385.832:621.317.7

Tuning indicators ("magic eyes") are employed in radio receivers and also, in bridge and balance circuits, as null-indicators for voltage comparison. The tuning indicator tubes hitherto available have the disadvantage that, generally speaking, they only show whether a voltage difference is present or not; they do not give its sign and only give a very rough idea of its magnitude.

These shortcomings have been eliminated in the E 82 M tuning indicator tube, developed in the laboratory of the Valvo radio tube factory in Hamburg<sup>1</sup>). In this tube an electron beam of rectangular section produces a rectangular spot on a fluorescent screen. Unlike other indicator tubes, the E 82 M has two symmetrically mounted electrodes which act as deflector plates. The voltages to be compared are applied to these electrodes either directly or via amplification stages. If the two voltages are not of equal magnitude there is deflection of the spot to one side or the other; further, the width of the spot when not deflected (i.e. when the voltages are equal) gives some indication of the value of the two voltages.

In a tube like this there must be some sort of datum to mark the position of no deflection, i.e. the central zero. The datum for the E 82 M is on the outside of the tube, in order to facilitate zero calibration, in connection with possible departures from symmetry in the electrode system or the circuit. Another reason for having the marker on the outside is to make the tube better adaptable to purposes which may require special forms of indication. The external datum mark necessitates a change in the arrangement of the fluorescent screen: this is viewed, not from the side on which the electrons strike it (as in the conventional tuning indicator), but from the side opposite to the beam. The screen is in fact coated on the inside wall of the tube, to avoid parallax errors. The datum mark may take the form of a pattern painted on the outside of the tube; alternatively, it may be a thin metal mask, cut out in any desired pattern, fitted on to the cylindrical envelope. One pattern that is suitable for many purposes appears in *fig. 1*; *fig. 2a* shows the complete tube, and *fig. 2b* a tube with the mask holder removed and four alternative mask patterns (numbered 1-4).

Fluorescent screens on glass are nothing new; they have long been used in oscillograph and television picture tubes. In such tubes the electron accelerating potential is of the order of several thousand volts. A tuning indicator, however, must

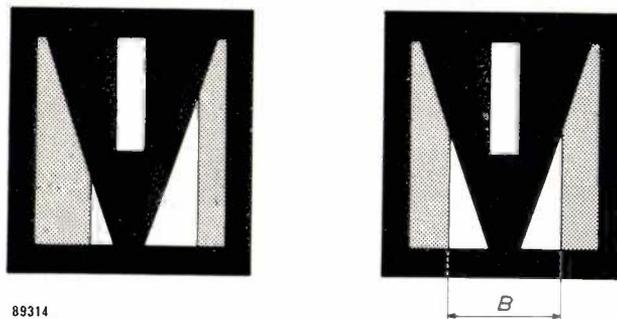


Fig. 1. Mask for the E 82 M tuning indicator. Left: appearance when tuning or adjustment is out of balance. Right: appearance when balanced. The breadth  $B$  of the spot (and hence also the height of the luminous triangles, when balanced) is a relative measure of the mean value of the two applied voltages.

rely on the supply already available in the receiver or measuring set; thus the acceleration voltage should not be higher than, say, 250 V. At voltages as low as this the normal type of fluorescent screen acquires a considerable charge, so that further electrons are prevented from striking it, and thus it remains dark; at higher voltages, increasing secondary emission drains off the charge quickly enough for this phenomenon to be avoided<sup>2</sup>). Where the screen is on a metal base, as in conventional tuning indicators (where, it will be remembered, the screen is viewed from the electron beam side, not by transmission as in the present tube) this base serves to conduct the charge away, even at low voltages. In the present tube the charge is drained off by means of a very thin layer of tin oxide coated on the inside wall prior to the application of the fluorescent screen; the tin oxide is practically transparent, but it has good conducting properties, and it is to this layer that the positive acceleration potential is applied. In this way the voltage below which charge would accumulate on the screen has been reduced to about 100 V. The phosphor used is zinc oxide, which fluoresces with a bluish-green light and which, after further activation by electron bombardment, exhibits a slower decrease of brightness with age than the usual willemite.

<sup>1</sup>) See also H. te Gude and E. Schaaff, *Abstimmanzeigeröhren für die Messtechnik, Elektronische Rundschau* 9, 184-189, 1955.

<sup>2</sup>) For a detailed discussion of these phenomena, see for example, J. de Gier, A. C. Kleisma and J. Peper, *Secondary emission from the screen of a picture-tube*, *Philips tech. Rev.* 16, 26-32, 1954/55.

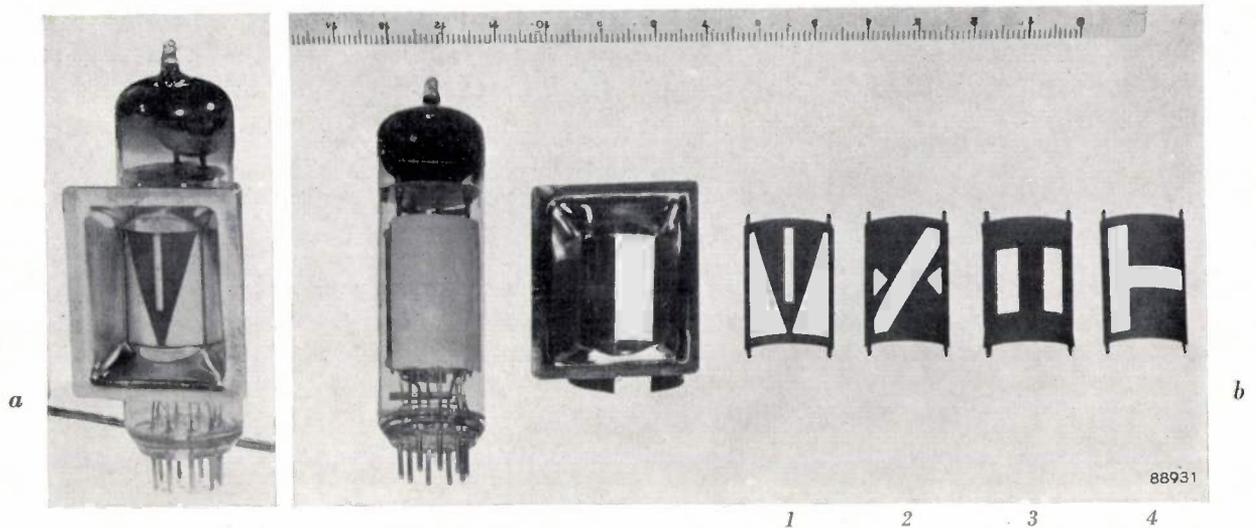


Fig. 2. a) E 82 M indicator tube fitted with mask in holder. b) Tube, mask holder and (1-4) various types of masks.

In order that the tube should have greater deflection sensitivity, a stage of amplification for each of the two voltages to be compared is incorporated within the same envelope. Each of the voltages is applied to the grid of a triode that acts as a voltage amplifier; the anodes of these two triode systems at the same time function as deflector plates for the indicator section of the tube. The actual electrode arrangement is shown in fig. 3.

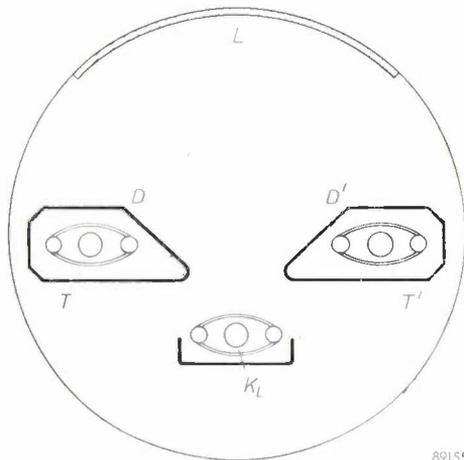


Fig. 3. Cross-section of the E 82 M tube.  $K_L$ : cathode of indicator section providing electrons for rectangular beam.  $L$  fluorescent screen.  $D$  and  $D'$ : deflector plates, which also form the anodes of the triode sections  $T$  and  $T'$ .

The characteristics of the two triode systems must, of course, be exactly the same within the operating range; differences of slope due to changes in the working point would have the same effect as a difference between the voltages to be compared. To ensure exact symmetry in this sense, the charac-

teristics of the triode sections of the E 82 M are made as straight as possible. It would hardly be possible to manufacture the tube without troublesome departures from symmetry if its characteristics were curved, as they are in many conventional tuning indicators, which have variable- $\mu$  characteristics in order to extend the range of voltages that may be compared. The E 82 M can also be given a wide measuring range, however, in spite of its straight-line characteristics, and this is achieved in the following way. The two triode systems are each provided with a separate anode resistor, while they share a common cathode resistor. The cathode resistor and the cathode bias voltage are both relatively large so that the negative bias then effective on the grid of each triode changes in the same sense and by approximately the same amount as changes in the mean value of the two voltages to be compared. This produces a beam-width versus mean-voltage characteristic, the working region of which, though roughly straight, has a smaller slope than it would have in the absence of the cathode resistor. The beam width thus gives a less sensitive indication of the mean value of the two applied grid voltages, which means that the operating range is increased and moreover this range can be varied during operation. Whatever the magnitude (within certain limits) of the applied voltages, their difference is always indicated in full. In order to permit the range adjustment facility for the mean voltage, the cathodes of the two triode sections are taken to a separate pin from that of the cathode of the indicator section.

Apart from its obvious uses, particularly as a null indicator in bridge circuits, where indication

of the sign of the unbalance is of great value especially in the choice of the correct measuring range, the E 82 M can be used for a number of different purposes. The two triodes can be made to function as a multivibrator in a flip-flop (scale of two) counter circuit, a suitable mask pattern (3 in fig. 2) being fitted to give indication of the counting state. Fitted with another mask pattern such as 2 or 4 in fig. 2, the tube becomes suitable (as are conventional tuning indicators) for giving a check

on modulation depth in tape-recorders and the like. Mention may also be made of the use of the tube (with mask 1 in fig. 2) as a tuning indicator for FM receivers, and for voltage measurement with a suppressed zero level (alarm circuits set off by the deviation of the unknown voltage from a standard comparison voltage).

In these and other applications it is sometimes a great advantage that the E 82 M may be operated entirely on alternating voltages. Periodical variation of the voltage on the deflector electrodes does no harm, provided only that this voltage remains proportional to that on the fluorescent screen, and this is automatically the case when the amplifying sections are fed from the same alternating voltage as the fluorescent screen anode. It should be noted that when using the range adjustment facility described above, the bias voltage on the common cathode resistor  $R_k$  must be in opposite phase to the voltage on the screen (see fig. 4).

H. te GUDE\*) and E. SCHAAFF\*).

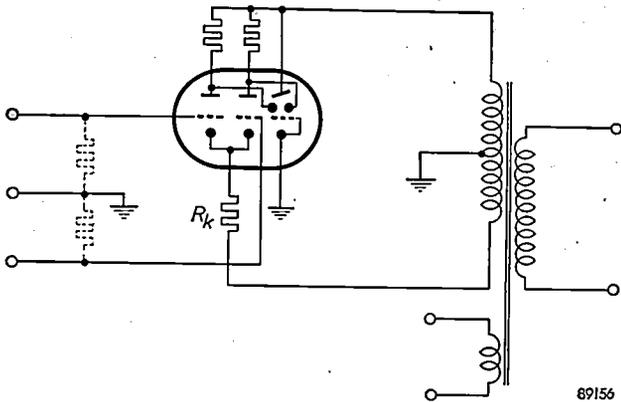


Fig. 4. Null-balance indicating circuit using the E 82 M, the tube being operated entirely on A.C.

\*) Development laboratory, Valvo G.m.b.H. Radioröhrenfabrik, Hamburg.

## ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk \* can be obtained free of charge upon application to the Philips Research Laboratory, Eindhoven, Netherlands.

- 2357:** R. Vermeulen: Music reproduction (T. Ned. Radiogenootschap 21, 39-41, 1956).

Short description of subjective experiments on the reproduction of music and the use of artificial stereo-reverberation (see also Philips tech. Rev. 17, 171-177 and 258-266, 1955/56).

- 2358:** G. Diemer and W. Hoogenstraaten: Evidence for hole-mobility in CdS (Physica 22, 172, 1956).

Measurements of the photoconduction under local illumination have been carried out on single crystals of CdS. When illuminated between graphite electrodes by a narrow line of ultraviolet (3650 Å, width about 0.5 mm, i.e.  $\frac{1}{20}$  of inter-electrode distance), these crystals show an increase of the current, at 250 volts D.C., by a factor of 20. This increase of the conductivity can be explained only if the space charge of the conducting free electrons in the dark part of the crystal is compensated by a nearly equal quantity of positive holes, which must therefore be assumed to migrate from the illuminated part through the whole inter-electrode space.

- 2359:** D. de Nobel and D. Hofman: The dielectric constant of CdTe (Physica 22, 252, 1956).

The dielectric constant of CdTe was determined at 20 °K and 77 °K at frequencies varying from 1 to 100 kc/s with a bridge of the Schering type. The dielectric constant is found to be approximately 11.

- 2360\*:** J. Volger, J. M. Stevels and C. van Amerongen: Les pertes diélectriques de divers "monocristaux" de quartz aux très basses températures (Verres et Réfractaires 10, 3-14, 1956).

Translation into French of R 272.

- 2361:** A. van Weel: Récepteurs de télévision à phase linéaire (Onde Electrique 36, 48-56, 1956).

The author shows that it is possible, without complicated circuitry, to design an I.F. amplifier with only a very small phase distortion yet fulfilling the normal requirements for television receivers. The same subject is dealt with in Philips Res. Rep. (see R 273 of these abstracts).

- 2362:** J. Smit: Galvanomagnetic properties of ferromagnetic metals and alloys (Dissertation, Leiden, 11 April 1956).

A number of measurements and theoretical work by the author are here collected together. In chapter I the band structure of incomplete shell electrons 3d and 4s in nickel is calculated, taking into account correlation. Chapter 2 gives experimental results for the Hall effect in ferromagnetic metals and alloys and a theoretical discussion on the origin of the spontaneous part. It is shown that in a perfectly periodic lattice the spontaneous Hall effect cannot occur. The explanation is to be found in anisotropic scattering of the conduction electrons at lattice faults, due to spin-orbit interaction. This is worked out numerically with the aid of a simple model. Chapter 3 deals with the magnetoresistance of ferromagnetics. Resistance anisotropy is explained on the basis of ellipsoidal charge distribution of the core electrons caused by spin-orbit interaction. This is discussed quantitatively (see also these abstracts Nos. 1996 and 2319).

- 2363:** N. W. H. Addink: Bestimmung von Spurenelementen (ohne chemische Vorbehandlung) und Analyse kleinster Materialmengen mittels spektrochemischer Methoden (Mikrochimica Acta, No. 1-3, 1956, pp. 299-303). (Estimation of trace elements, without chemical pre-treatment, and analysis of very small amounts of material by spectrochemical methods; in German.)

The lowest limits of detection of several trace elements are given for spectrochemical and microchemical determinations. There is a relationship between the lowest limits of detection as determined by spectrochemical procedures at various temperatures. Some figures regarding the minimal amount of the sample are given.

- 2364:** A. L. Stuijts: Sintering of ceramic permanent magnetic material (Trans. Brit. Cer. Soc. 55, 57-74, 1956).

The permanent magnetic properties of the hexagonal compound  $\text{BaFe}_{12}\text{O}_{19}$  are described. To achieve the best magnetic values the product should have a high apparent density and consist of particles with a diameter of the order of 1  $\mu$ . Sintering of a non-

stoichiometric composition is favourable for obtaining high densities. In order to keep the crystallites as small as possible, a high rate of heating up and a short time at the highest temperature are desirable.

In sintering, two different types of crystal growth are observed: a continuous and a discontinuous one. In the latter case, an explanation may be found in the presence of a second phase. A considerable improvement in magnetic properties can be obtained by lining up the hexagonal axes of the component crystallites. In practice this can be achieved by pressing a paste of the powder in a magnetic field. Crystal growth during sintering enhances the magnetic anisotropy of the material considerably.

**2365:** J. S. C. Wessels and R. van der Veen: The action of some derivatives of phenylurethan and of 3-phenyl-1,1-dimethylurea on the Hill reaction (*Biochim. et biophys. Acta* **19**, 548-549, 1956).

The action of some derivatives of phenylurethan and of 3-phenyl-1,1-dimethylurea on the Hill reaction was investigated. 3-phenyl-1,1-dimethylurea derivatives inhibit the Hill reaction at lower concentrations than derivatives of phenylurethan. The inhibitory activity is enhanced by introduction of electron-attracting groups into the benzene nucleus. A possible explanation of the experimental results is presented.

**2366:** J. J. Balder: Verfärbung farbiger Objekte durch die Einwirkung von Tageslicht, Glühlampenlicht und Leuchtstofflampenlicht (*Lichttechnik* **8**, 57-81, 1956). (Discoloration of coloured objects under the influence of daylight, incandescent light and fluorescent light; in German.)

The discoloration of a number of coloured objects by various types of light was investigated. Four specimens of each object were exposed to four types of light: daylight filtered through window glass, incandescent lamp light, and light from tubular fluorescent lamps "White-de-luxe" and "Warm-white-de-luxe". The dose of light administered was in all cases 8 Mluxh. The discolorations were judged visually by three persons. This method could be justified by colorimetric measurements. Daylight through window glass had the worst discolouring effect, followed by fluorescent lighting White-de-luxe and then incandescent lighting, while fluorescent lamps type Warm-white-de-luxe gave least discoloration.

**2367\*:** J. Hermsen, A. M. J. Jaspers, P. Kraayeveld and K. van Duuren: New Geiger tube

designs, hollow anode and parallel plate counters. (*Proc. Int. Conf. Peaceful uses Atomic Energy, Geneva, Aug. 1955, Vol. 14*, pp. 275-276, New York 1956).

Short notice concerning some Geiger counters of unusual geometry. One type has a tubular anode (internal diameter 23 mm), in place of the usual wire; the sample may be placed within this tube. Another type has parallel plate electrodes. The properties of these counters are given in the form of graphs.

**2368:** J. A. Kok: Experiments with gas-filled triodes (*Appl. sci. Res. B* **5**, 445-453, 1956).

In this paper experiments are described showing the different types of electrical discharges in a gas-filled triode. The determining parameters are the following: the cathode emission, the spacing of cathode, grid and anode, the diameter of the meshes of the grid, the potentials of the grid and the anode; the gas pressure and the differential ionization function of the gas. The anode voltage may be concentrated in a space-charge sheath. If this space-charge sheath is located at the grid, the anode current may be modulated with moderate grid potentials. If not, much larger voltages are required for modulation.

**2369:** K. H. Klaassens and C. J. Schoot: Some fluoro-substituted phenoxyacetic acids (*Rec. trav. chim. Pays-Bas* **75**, 186-189, 1956).

The preparation of 2,4-dichloro-6-fluorophenoxyacetic acid, 2,4-dichloro-3-fluorophenoxyacetic acid and 2,4-dichloro-5-fluorophenoxyacetic acid is described.

**2370:** C. J. Schoot and K. H. Klaassens: 5,7-dichlorocoumaran-3-one (*Rec. trav. chim. Pays-Bas* **75**, 190-192, 1956).

The preparation of 5,7-dichlorocoumaran-3-one is described.

**2371:** J. F. Klinkhaimer: Reflectionless transmission through  $2n$ -terminal-pair networks (*Proc. Symp. Mod. Network Synthesis, Polytechnic Inst. Brooklyn, April 1955*).

Various authors have treated  $2n$ -poles by a generalization of corresponding considerations of 2-poles. In such cases current and voltage vectors take the place of the scalar currents and voltages of the 2-pole case, while impedance matrices take the place of the normal scalar impedances. In a similar way, in this article, by a generalization of considerations of a 4-pole chain, a "mirror-impedance" and a "trans-

mission constant" are defined in matrix form for a chain of  $4n$ -poles. Both matrices, for positive values of the real part of the frequency parameter, are unambiguously determined. The eigen vectors of the transmission matrix can be considered as independent travelling waves, each with its own damping and phase shift.

**2372:** C. Z. van Doorn and D. de Nobel: Luminescence, transmission and width of the energy gap of CdTe single crystals (*Physica* **22**, 338-342, 1956).

Irradiation of a CdTe single crystal or biasing a CdTe p-n junction in the forward direction lead to a luminescence with a maximum at about 8880 Å. From transmission measurements and measurements of the long wave limit of the photo-e.m.f. of a CdTe p-n junction as a function of the temperature, the shift of the band edge was seen to vary from  $2.34 \times 10^{-4}$  eV/°K at 77 °K to  $5.44 \times 10^{-4}$  eV/°K at 800 °K.

**2373:** J. L. Ouweltjes and W. L. Wanmaker: Contributions to the problem of nonstoichiometry in oxygen-dominated phosphors (*J. Electrochem. Soc.* **103**, 160-165, 1956).

According to general experience, oxygen-dominated phosphors of highest efficiencies are obtained with non-stoichiometric compositions. This has been explained by the hypothesis that crystal defects are essential to obtain luminescence. In this paper it is postulated that the ideal, and therefore stoichiometric crystal should have maximum efficiency. That an excess of one of the ingredients in general results in higher efficiency, is explained with certain factors arising from the mechanism of the solid-state reactions, such as the difficulty of reaching equilibrium conditions. The incorporation of the activator, the presence of U.V. absorbing separate phases, the problem of homogeneity of the phosphors, and the reactivity of the ingredients used for the synthesis are discussed, mainly on the basis of

experimental evidence gathered in the study of the phosphors used in fluorescent lamps.

**R 298:** P. Zalm: The electroluminescence of ZnS type phosphors (*Philips Res. Rep.* **11**, 353-399, 1956).

First part of a paper dealing with various aspects of the electroluminescence of ZnS (Destriau effect). Section I gives a brief survey of the theories of photoluminescence in ZnS, in so far as these are of consequence for electroluminescence. Various methods of preparation of electroluminescent ZnS powders are treated in section 2. Section 3 deals with electrical and optical measurements which have led to a qualitative model with which the mechanism of electroluminescence can be explained. A distinction must be made between phosphors where excitation by the field causes ionization of the activators, and those where this does not happen. In the latter case the light emission is virtually in phase with the voltage; in the former case, electrical factors cause a shift over a substantial fraction of the period of the field between the moments of maximum excitation and of maximum emission. In experiments with hexagonal single ZnS crystals activated by Cu it was found that the mechanism of electroluminescence does not differ essentially from that of powders. In single crystals the excitation occurs at the internal barriers, which have rectifying properties. A relationship is established between the orientation of the barriers at which light emission occurs and that of the crystal axes. For the second part of this paper, see **R 300**.

**R 299:** M. J. Sparnaay: On thermal hysteresis of transformations in solids (*Philips Res. Rep.* **11**, 400-409, 1956).

A simple order-disorder theory for the qualitative understanding of hysteresis phenomena is discussed. The essential feature is the introduction of an element of asymmetry into a zeroth-order treatment.

# Philips Technical Review

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

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

## COLOUR AND COLOUR RENDERING OF TUBULAR FLUORESCENT LAMPS

by A. A. KRUIHOF and J. L. OUWELTJES.

621.327.534.15:535.625

---

*The tubular fluorescent lamp has given rise to all kinds of new problems in lighting technology which were irrelevant or insoluble in the case of the incandescent lamp. Illumination by distributed light sources (as opposed to point sources) required new investigations; the design of fittings again raised the question as to permissible brightnesses; the high luminous efficiencies and the white colour of the light from the first fluorescent lamps led to reconsideration of desirable levels of illumination; the fact that it was possible to vary the spectral distribution of the light from the lamps within wide limits, led to investigations regarding the most desirable colour for the light and discrepancies in the colour rendering. This article describes how the latter investigations have resulted in a range of lamps with a largely natural colour rendering.*

---

### The colour of tubular fluorescent lamps

It was about 25 years ago that the first attempts were made to coat the wall of the low-pressure mercury-vapour lamp with a fluorescent powder (phosphor) and in this way to arrive at a lamp for general lighting purposes. The phosphor converts the ultra-violet radiation from the low-pressure mercury-vapour discharge, especially the spectrum line of wavelength 2537 Å, into visible light. This conversion can be brought about with high yield, so that the resulting lamp may have a very high luminous efficiency, much higher than that of the incandescent lamp. High luminous efficiency is one of the chief features of the modern tubular fluorescent lamp, which is the fruit of those early efforts.

The search for suitable fluorescent substances was soon so successful that it became possible by using a mixture of phosphors in the fluorescent coating to produce light in a wide variety of colours. The question then arose as to exactly what colour was desirable.

Initially the only guiding principle was that lamps for general lighting purposes should give an impression of "whiteness". Closer study reveals that "white" is a far from well-defined colour. The human eye is willing to see a white tablecloth as "white" both by daylight and by incandescent-lamp light,

and, indeed, even by candlelight, despite the fact that the radiation it receives from the cloth is totally different in each case. This is known as "chromatic adaptation" of the eye<sup>1</sup>).

The differences between the radiations from the above-mentioned sources are only apparent to the eye if they are present in the visual field at one and the same time. Now this state of affairs was frequently encountered in fluorescent-lighting practice. For example, fluorescent lamps had to be used for supplementary day-time lighting in dark shops and offices, or in combination with incandescent lighting which was already present, or was required for special effects. When colour differences were too great, false effects and coloured shadows resulted — troublesome phenomena which had already been met with, of course, in combining incandescent-lamp light with daylight. They had then been accepted as unavoidable in most cases. With fluorescent lamps, where the choice of fluorescent substances offered greater freedom, there was no need to accept these troublesome phenomena. Hence the vague guiding principle, already referred to, was subsequently narrowed down to state that the colour

---

<sup>1</sup>) A. A. Kruithof and P. J. Bouma, *Physica* 9, 957-966, 1942; P. J. Bouma and A. A. Kruithof, *Physica* 10, 36-46, 1943. See also *Philips tech. Rev.* 9, 2-7 and 257-266, 1947/48.

of the fluorescent light must resemble that of daylight or that of incandescent-lamp light. A complication that now arises is that "daylight" can vary continuously within wide limits according to the height of the sun and to the weather conditions. The colour of incandescent-lamp light can also vary somewhat according to the temperature of the filament. In order to fix the desired colour more exactly, use is made of the fact that most variants of daylight, as well as incandescent-lamp light, closely resemble the radiation emitted by a black body at a certain temperature. Hence these sources can be characterized by the corresponding temperature of the black body (colour temperature). *Table I* gives the colour temperatures for a number of light sources.

Table I. Colour temperatures for a number of light sources.

Light source	Colour temperature in °K
Blue sky	10000-20000
Overcast sky	5000-7500
Direct sunlight (midday)	approx. 5000
Carbon arc	approx. 3750
Photoflood lamp	approx. 3200
Incandescent lamp	2400-3000
Candle	1900

In addition to fluorescent lamps for combination with daylight and others for combination with incandescent lamps, lamps which might be used with both types of light were considered. Naturally such lamps involve a compromise. In order to explain the choice of the colour for such a lamp, we must make use of the well-known representation of colours by points in a chromaticity diagram<sup>2)</sup>; see *fig. 1*. The points for the incandescent black body at various temperatures lie on a certain curve in the figure — the black-body locus — and on or very near this curve will be situated the points for the selected types of "daylight" (D) and "incandescent-lamp light" (G). An obvious choice for the compromise would be the colour point which lies halfway along the line connecting D and G. However, the colour contrasts with respect to D and G are found to be somewhat less troublesome for a colour which is situated on (or very close to) the black-body locus and whose colour temperature is nearer that of G than that of D.

<sup>2)</sup> See in this connection P. J. Bouma, Philips tech. Rev. 1, 283-287, 1936, and 2, 39-46, 1937; a more comprehensive exposition is to be found in the book by the same author: Physical aspects of colour, Philips Technical Library, 1946.

Another consideration in designing a lamp with its colour point between D and G, was that the colour Daylight was only found to be agreeable at relatively high levels of illumination — even when the lamp was not used in combination with natural daylight. This point has been discussed at length in an earlier article in this journal<sup>3)</sup>.

Thus by a process of trial and error, three standard types of fluorescent lamps have been arrived at, which conform to the practical requirements and have the following colour temperatures and colour points<sup>4)</sup>:

- "Cool Daylight", 6800 °K, with trichromatic coefficients (coordinates in the chromaticity diagram)  $x = 0.308$ ,  $y = 0.324$  (see *fig. 1*). (This lamp has no equivalent in England, but corresponds to the "Daylight" lamps of American manufacturers.) Lamps of this type are suited for use in combination with average daylight.
- "Warm White", 2900 °K, trichromatic coefficients  $x = 0.440$ ,  $y = 0.399$ . These lamps are suitable for use alongside incandescent lamps; their light is only slightly bluer than that of the larger incandescent lamps (colour temperature between 2600 and 2850 °K).
- "White", 4200 °K, trichromatic coefficients  $x = 0.371$ ,  $y = 0.368$ . If necessary, these lamps can be combined both with daylight and with incandescent lamps. (This Dutch lamp corresponds to the English "Daylight" lamp ( $T_c = 4500$  °K) and to the American "Cool White" ( $T_c = 4400$  °K).)

In order to be able to make these lamps, it is necessary to have available phosphors which give sufficient latitude to allow the desired colour points to be realized. How great this latitude is with three current phosphors can easily be seen if the colour points of these phosphors are indicated in the chromaticity diagram. In *fig. 1*, the colour points are shown for three phosphors which are suitable for use in fluorescent lamps, viz. calcium silicate, activated with lead and manganese; calcium tungstate; and zinc silicate activated with manganese (willemite). Any colour point within the triangle formed by these three points can be obtained by mixing the three phosphors mentioned in the correct proportions. Nowadays halophosphates are often used for the fluorescent coating. With these materials there is a certain amount of colour mixing in the phosphor itself<sup>5)</sup>: halophosphates have two

<sup>3)</sup> Philips tech. Rev. 6, 69, 1941, *fig. 10*.

<sup>4)</sup> Attempts are also being made to gain international acceptance of three standard colours, whose colour points differ only slightly from those reported here.

<sup>5)</sup> J. L. Ouweltjes, New phosphors for fluorescent lamps, Philips tech. Rev. 13, 346-351, 1951/52.

emission bands, one in the blue (at about 4800 Å) and one in the orange (at about 5900 Å). The location of the latter band and the ratio between the intensities of the two bands can be varied somewhat by varying the composition of the halophosphate, so that here also a diagram of colour points, albeit a narrow one, can be obtained.

One important advantage of the halophosphates is that the colour points of the lamps are more readily reproducible in mass production. Also, small deviations from the correct composition give rise to a displacement of the colour point approximately along the black-body locus. Such a displacement is far less troublesome than a displacement at right angles to this line.

To conclude this introduction we shall describe briefly how the proportions are calculated in which three given phosphors (or, rather, their light contributions) must be mixed in order to produce a lamp with a predetermined colour point ( $x, y$ ). The formulae used in this calculation form the basis of all the calculations which will be discussed later in this article.

The formulae are simplest if we introduce the trichromatic coordinates (coordinates in the three-dimensional colour space) of the colour points of the phosphors:  $X_1, Y_1, Z_1; X_2, Y_2, Z_2; X_3, Y_3, Z_3$  <sup>6)</sup>. Each trichromatic coordinate of the colour of a

<sup>6)</sup> For fuller details regarding the coordinates  $X, Y, Z$  in colour space and the trichromatic coefficients  $x, y$  which fix the location of a point in the chromaticity diagram, see for example the book cited in <sup>2)</sup>, pp. 73 et seq. and 80 et seq.

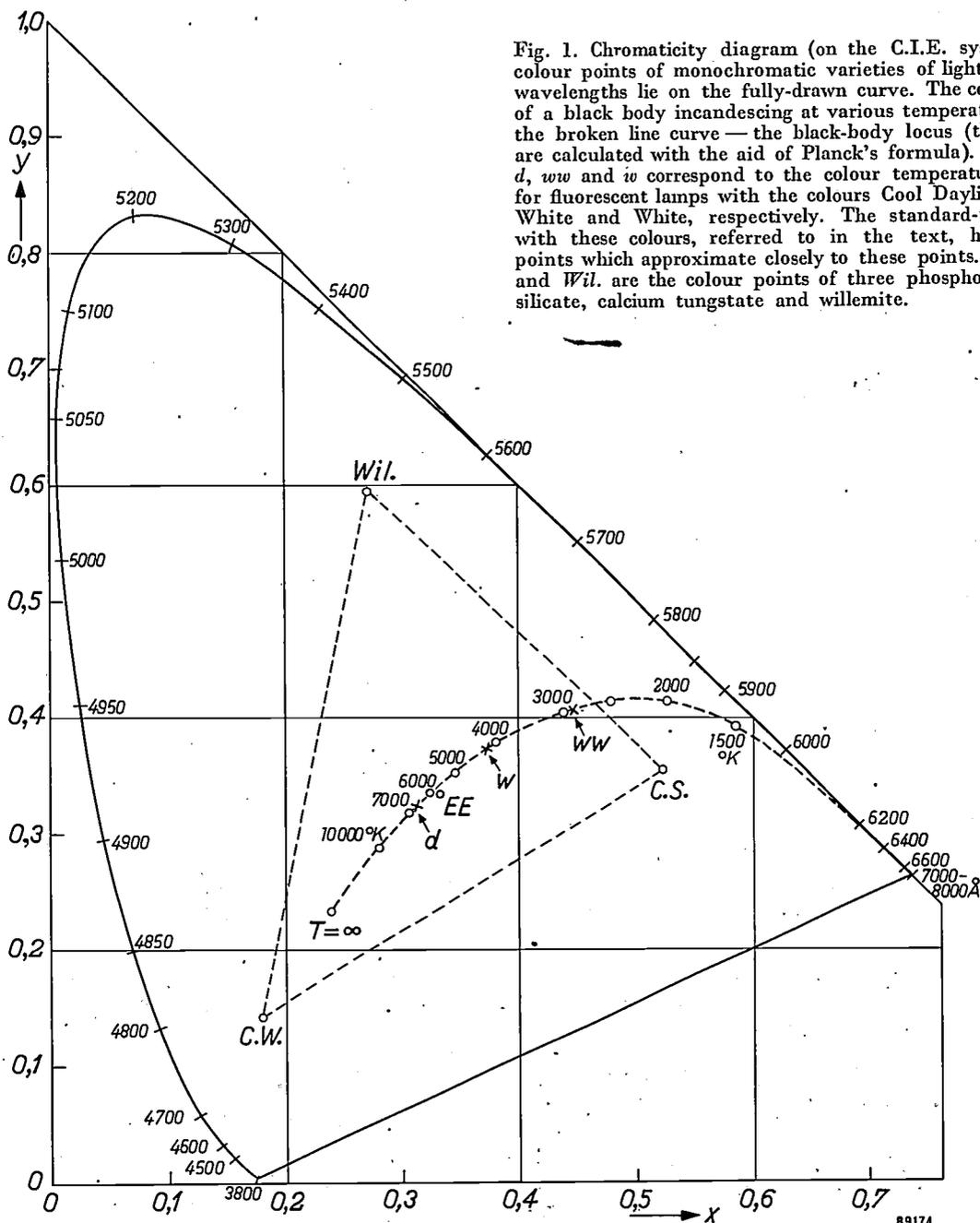


Fig. 1. Chromaticity diagram (on the C.I.E. system). The colour points of monochromatic varieties of light of various wavelengths lie on the fully-drawn curve. The colour points of a black body incandescing at various temperatures lie on the broken line curve — the black-body locus (these points are calculated with the aid of Planck's formula). The points  $d, ww$  and  $w$  correspond to the colour temperatures chosen for fluorescent lamps with the colours Cool Daylight, Warm White and White, respectively. The standard-type lamps with these colours, referred to in the text, have colour points which approximate closely to these points. *C.S.*, *C.W.* and *Wil.* are the colour points of three phosphors: calcium silicate, calcium tungstate and willemite.

mixture is the sum of the respective coordinates of the components, each multiplied by a factor ( $\alpha, \beta, \gamma$ ) which determines the mixing proportion. The coordinates  $Y_1, Y_2, Y_3$  indicate the value of the luminous flux which each of the phosphors would supply in a 40-watt lamp containing that phosphor only (all calculations are based on a constant lamp power, viz. 40 W). In a 40-watt lamp the mixture then gives a luminous flux of  $\alpha Y_1 + \beta Y_2 + \gamma Y_3$ . The luminous fluxes  $Y_1, Y_2, Y_3$  are of course different, since the phosphors convert the absorbed ultra-violet quanta into quanta to which the human eye is sensitive to varying degrees. The "quantum yield", i.e. the number of visible quanta emitted divided by the number of ultra-violet quanta absorbed, is almost the same for all phosphors in practical use, being about 90%; hereafter we shall assume that it is always exactly the same.

Apart from fluorescent light, a fluorescent lamp always emits a certain amount of visible light that originates directly from the mercury vapour discharge; this contribution to the luminous flux is absorbed hardly at all by the normal fluorescent layer and depends therefore only on the given power of the lamp and not on the nature of the phosphors or the proportions in which they are mixed. The trichromatic coordinates of the colour point for this mercury light we shall call  $X_{Hg}, Y_{Hg}, Z_{Hg}$ , those of the light from the resulting lamp  $X, Y, Z$ . We then have:

$$\left. \begin{aligned} X &= \alpha X_1 + \beta X_2 + \gamma X_3 + X_{Hg}, \\ Y &= \alpha Y_1 + \beta Y_2 + \gamma Y_3 + Y_{Hg}, \\ Z &= \alpha Z_1 + \beta Z_2 + \gamma Z_3 + Z_{Hg}. \end{aligned} \right\} \dots (1)$$

The factors  $\alpha, \beta, \gamma$  must satisfy the condition

$$\alpha + \beta + \gamma = 1, \dots (2)$$

since the quantum yield is the same for all the phosphors and the total number of quanta available to the phosphors for conversion is constant for a given discharge wattage (i.e. it is independent of the composition of the fluorescent coating); increase of any one contribution therefore is always at the expense of the other contributions.

Further, for the desired lamp the following trichromatic coefficients<sup>6)</sup> are stipulated:

$$\left. \begin{aligned} x &= \frac{X}{X + Y + Z}, \\ y &= \frac{Y}{X + Y + Z}. \end{aligned} \right\} \dots (3)$$

(The reason why it is impossible to stipulate the three trichromatic coordinates of the lamp is connected with the fact that the total luminous flux of the lamp for a given power cannot be predicted, since the resultant luminous efficiency of the phosphor mixture is for the time being unknown.) The six quantities  $\alpha, \beta, \gamma, X, Y, Z$  are determined by the six equations (1), (2) and (3). Thus the required proportions in which the phosphors must be mixed can be calculated. If the stipulated colour point lies within the above-mentioned triangle of the colour points of the phosphors, positive values are always found for  $\alpha, \beta, \gamma$ , i.e. the colour can be realized.

Colour rendering

Thus far we have only discussed colour points, but have ignored the fact that sources with quite different spectral energy distributions may have the same colour point. This is illustrated by fig. 2. Radiation  $a$  with a so-called equal energy spectrum (the energy in a narrow wavelength range  $\Delta\lambda$  is the same at all wavelengths) appears "white" and has the colour point  $x = 1/3, y = 1/3$  ( $EE$  in fig. 1). Radiation  $b$  comprising only the wavelengths  $\lambda = 5890 \text{ \AA}$  and  $\lambda = 4860 \text{ \AA}$ , the latter having twice the energy of the former (both for the sake of clarity broadened out in the figure into a wavelength range of  $200 \text{ \AA}$ ), has the same colour point and hence gives the same impression of whiteness. The colours of objects, however, are in general rendered quite differently by the two types of light. A red object,

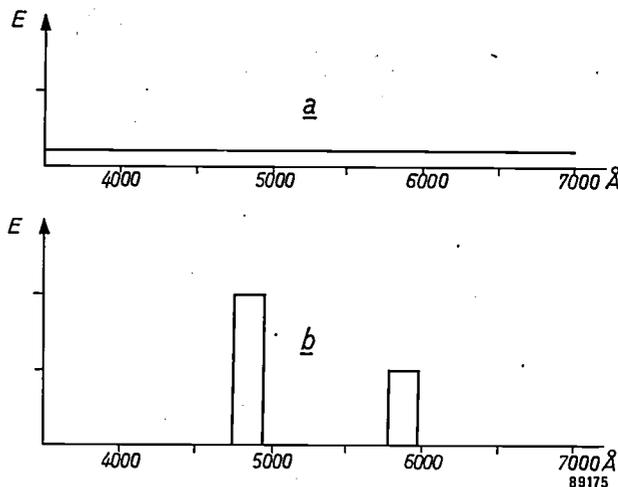


Fig. 2. The light source with the spectral energy distribution  $a$  (equal energy spectrum) has the identical white colour of and the same luminous flux as that with the spectral distribution  $b$ . The colour point of both light sources in the chromaticity diagram in fig. 1 is  $EE$  ( $x = 1/3, y = 1/3$ ).

reflecting only the red rays of the light falling upon it, will appear black or dark brown by light  $b$ . By light  $a$  it will have its natural colour, and this will be the case not only for red objects but for objects of any colour. In light  $b$  only yellow and blue objects will retain approximately their natural colours. The colour rendering of  $b$  is therefore poor, that of  $a$ , on the other hand, very good or "natural".

Owing to the chromatic adaptation of the human eye, there are quite a number of types of light in which almost all objects give us the impression of having their natural colours, in spite of differences in the spectral energy distributions of these types of light and the resulting differences in the colour coordinates for the same given object. Notable examples of such types of light are daylight in all its varieties, incandescent-lamp light and,

we may assume, the radiation from black bodies at intermediate temperatures. For these reasons (and also because they are found naturally) all thermal radiators are called "natural light sources".

Returning to fluorescent lamps, the above discussion shows that the mere establishment of the colour points for the three standard lamps in no way establishes the spectral energy distribution for each of these lamps. The latitude which still remains, and which therefore permits of choosing phosphors with a great diversity of spectral distributions as the components of the mixture, can be made use of to achieve two different ends:

- 1) to give the lamps the highest attainable luminous efficiency<sup>7)</sup>;
- 2) to make the colour rendering of the lamps as good as possible.

In the above-mentioned standard-type lamps, high luminous efficiency is given first consideration; for the Cool Daylight lamp, a luminous efficiency of 55 lumens per watt has been attained, for the Warm White lamp 67.5 lm/W and for the White lamp 61 lm/W. (Incandescent lamps for general lighting give 20 lm/W at most.) In addition the past few years have witnessed intensive work on fluorescent lamps in which the emphasis has been in the other direction — on the colour rendering qualities. The colour points of such lamps should be practically the same as before. In what follows we shall see how the task of improving the colour rendering has been set about and what results have been obtained in regard to these so-called "de Luxe" lamps.

It is first necessary to show in which respects the colour rendering of the standard fluorescent lamps falls short. This can be done by comparing the spectral distribution of these lamps with that of natural light sources, i.e. black bodies with corresponding colour temperatures. This comparison is made in *fig. 3a-c* for the energy distributions and in *Table II* for the light flux, employing a subdivision of the spectrum into the eight wavelength bands introduced by Bouma<sup>8)</sup>. Further, there must be some means of assessing how badly the differences in the spectral distributions of the lamps affect their colour rendering qualities. A means of

assessment which was found empirically is Harrison's figure of merit. The figures of merit are calculated from the eight-band light-flux distribution<sup>9)</sup> and have been included in the table. A more direct, though more laborious method consists in directly

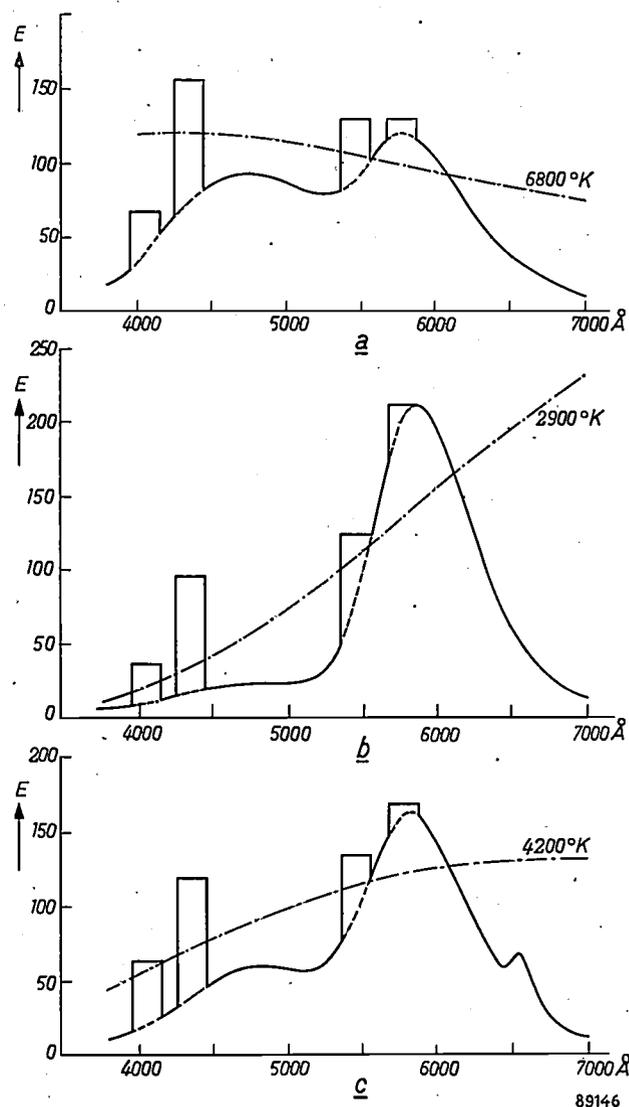


Fig. 3. The spectral energy distributions of the standard types of 40-watt fluorescent lamps (halophosphate phosphors).  
 a) Lamp colour "Cool Daylight", colour temperature 6800 °K, luminous flux 2200 lm.  
 b) "Warm White", 2900 °K, 2700 lm.  
 c) "White", 4200 °K, 2450 lm.

The energy scale for the fluorescent lamps is the same in all three figures. The broken line in each figure represents the spectral distribution of the corresponding black body, the scale being such that the luminous flux is the same as that of the lamp in question.

<sup>7)</sup> It will be clear that a high luminous efficiency is attainable if the spectral distribution of the components favours those spectral bands to which the eye is most sensitive. The halophosphates are especially rich in emissions of these wavelengths.  
<sup>8)</sup> P. J. Bouma, Philips tech. Rev. 2, 1-7, 1937. Later the subdividing was changed somewhat, so that the spectral bands shown in the table are now used. In 1948 this method of assessing colour rendering was recommended by the C.I.E. (Commission Internationale de l'Eclairage).  
<sup>9)</sup> Let  $l_i$  be the relative luminous-flux contribution of the lamp in the  $i^{\text{th}}$  spectral band,  $z_i$  that for the corresponding black body and  $E_i$  that for the radiator with an equal energy spectrum (fig. 2a). Harrison now calculates the ratios  $l_3/l_6, l_1/l_6, l_6/l_6$  and  $(l_7+l_8)/l_6$  and divides these by the corresponding ratios  $E_3/E_6$  etc. The average difference between the calculated ratios and the corresponding ratios for the black body (i.e.  $z_3/z_6 : E_3/E_6$  etc.) is multiplied by 100; subtracting the result from 100 gives Harrison's "figure of merit". See W. Harrison, Light and Lighting 44, 148-153, 1951.

comparing the colours of a series of objects in both kinds of light, either experimentally or by calculation. In the calculation the colour points of the test objects are determined when lighted with the two sorts of light to be compared, starting from the given spectral reflection curves (reflection coefficient as a function of the wavelength). Since the two sorts of light being compared have the same colour point (only their spectral energy distributions differ), the state of chromatic adaptation of the eye is the same, so that identity of colour point also signifies identity of colour sensation. Judd<sup>10)</sup> has constructed a chromaticity diagram (the Uniform Chromaticity Scale or UCS diagram) in which, given that the chromatic adaptation of the eye is constant, the distance between two colour points is approximately proportional to the difference in colour which is observed. A distance of 0.01 is

Table II. Relative spectral light-flux distributions of the standard fluorescent lamps and their corresponding black bodies. The spectrum is divided into eight bands after Bouma. Harrison's figure of merit  $H$  is also given for each kind of light.

Spectral band No.	Wave-lengths in Å	"Cool Daylight"		"Warm White"		"White"	
		Fluor- escent lamp	6800 °K	Fluor- escent lamp	2900 °K	Fluor- escent lamp	4200 °K
I	3800-4200	0.017	0.037	0.006	0.006	0.010	0.017
II	4200-4400	0.44	0.27	0.28	0.062	0.33	0.14
III	4400-4600	0.66	0.86	0.15	0.26	0.33	0.50
IV	4600-5100	8.9	10.9	2.3	5.5	5.1	8.1
V	5100-5600	41.8	42.0	29.2	33.7	35.7	38.6
VI	5600-6100	41.5	33.5	57.3	42.5	49.6	39.1
VII	6100-6600	6.5	9.7	10.5	16.4	8.8	12.5
VIII	6600-7800	0.15	0.67	0.21	1.54	0.21	0.99
Figure of merit $H$		67	100	62	100	64	100

clearly observable as a difference in colour. In the calculations for 15 coloured objects, some with saturated and some with unsaturated colours (one of the latter corresponding to the colour of the skin), the greatest colour discrepancy  $\Delta_m$  of the Cool Daylight lamp, compared with the black body at corresponding temperature, was found to be 0.045, that of the Warm White lamp 0.065 and that of the White lamp 0.05.

#### The development of the "de Luxe" colours

In attempts to improve the colour rendering qualities, the spectral energy distributions of the lamps, as given in fig. 3, were taken as a guide. The most serious deficiencies from the ideal distribution (i.e. that of the corresponding black body) are:

- 1) deficiency in the extreme red spectral region (spectral band VIII);
- 2) the occurrence of two humps in the spectrum, one in the blue-violet (mercury line 4358 Å) and one in the yellow-green (about 5800 Å). The two humps are interrelated: to compensate for the unavoidable strong blue-violet mercury line special provision has to be made in formulating the phosphor mixtures, to obtain a hump in the complementary colour, the yellow-green<sup>11)</sup>.

Qualitatively, it can readily be understood how the deviations from the ideal curve adversely affect the colour rendering. Owing to deficiency (1) (above) red colours are not faithfully reproduced: bright red colours assume a dull look. Deficiency (2) is particularly detrimental to the rendering of yellow and orange objects: the extra light contributions in the blue-violet and the yellow-green do, it is true, compensate each other on an object which reflects both kinds of radiation, e.g. a white cloth, but yellow and orange objects absorb the surplus blue-violet, so that the excess of yellow-green is no longer neutralized. The result is the familiar green discoloration which such objects assume when illuminated by standard fluorescent lamps. A notorious case is that of certain foodstuffs, such as butter, potatoes and fruit, which in this light often assume a somewhat unpalatable appearance. Similarly, blue-coloured objects are given a violet tinge, but this is usually found to be less troublesome.

Deficiency (1) is present in the same degree in all three types of fluorescent lamps (Cool Daylight, White, Warm White). Deficiency (2) demands more attention in the case of the Warm White and White lamps than in the case of the Cool Daylight lamp. This too is understandable. The contribution of the mercury light (and of the blue-violet mercury line, therefore) is pretty well independent of the colour of the lamp (see equation 1). In the Warm White and White lamps, which emit only a relatively small amount of blue radiation (figs. 3b and c), the latter is for the most part concentrated in the mercury line; in the Cool Daylight lamp for which the contribution of the continuous spectrum in the blue is much larger (fig. 3a), the contribution of the mercury line plays a less important role.

We shall now discuss in turn remedies that have been tried for both deficiencies.

<sup>11)</sup> The complementary colour of 4358 Å must be determined separately for each lamp by drawing a line from the point for that wavelength through the colour point of the lamp to the opposite side of the chromaticity diagram. With all three lamp colours discussed here we arrive at the yellow-green.

<sup>10)</sup> D. B. Judd, J. Opt. Soc. Amer. 25, 24-36, 1935.

*The admixing of red*

Starting from the standard-type lamps with a halophosphate fluorescent layer attempts have been made to remedy the deficiency in the red by adding a new, red-fluorescing phosphor to the phosphor mixture. Magnesium arsenate<sup>5)</sup> with a strong emission band in the deep red (maximum at 6560 Å) was found to be a suitable phosphor. When this phosphor is added to the mixture, the colour point of the lamp shifts in the first instance to the right and to below. In order to restore the colour point to its former position, the blue emission band of the halophosphate must be made stronger (shift to the left) and at the same time a green phosphor, e.g. willemite, must be added to give the necessary return shift upwards.

If for the sake of simplicity we regard the two bands of the halophosphate (orange and blue) to be derived from separate phosphors, we are then dealing with a mixture of four components. Obviously the mixing proportions are not now unequivocally determined by the assigned colour point: any one of the proportions can be given an arbitrary value. Here we have chosen the ratio between the two reddest phosphors. If we call this ratio  $a$ , we can then put the following question: for which value of  $a$  does the colour rendering of our mixture least deviate from that of the black body that corresponds to the lamp?

To be able to answer this question, we have again calculated the colour points in Judd's UCS diagram, of the fifteen coloured objects already mentioned, for a series of mixtures with various values of the parameter  $a$ , viz. 0, 0.2, 0.4, 0.5, 0.6, 0.8 and 1. In fig. 4 the calculated colour points  $L_p$  for five of the objects are indicated for  $a = 0$ ,  $a = 0.5$  and  $a = 1$ , for the lamp colour "White", which corresponds to that of a black body radiating at 4200 °K. The colour points  $N_p$  of the objects when illuminated with this black body are likewise given. The distance  $\Delta_p$  from  $L_p$  to  $N_p$  has been calculated for the above values of  $a$  and is plotted in fig. 5 as a function of  $a$ <sup>12)</sup> for all 15 objects ( $p = 1$  to 15).

Although the minima of  $\Delta_p$  do not occur for the same value of  $a$  for all the coloured objects, it can

<sup>12)</sup> The mixing of the two red phosphors with colour coordinates  $X_1, Y_1, Z_1$  and  $X_2, Y_2, Z_2$  in the ratio  $a$  means that we are replacing the two phosphors by one phosphor with coordinates  $aX_1 + (1-a)X_2, aY_1 + (1-a)Y_2, aZ_1 + (1-a)Z_2$ . The colour of an object  $L_p$  calculated for this mixture, however, can always be regarded as a mixture of the original colours as calculated for  $a=0$  and for  $a=1$ . It follows from this that, on variation of  $a$ , each colour point  $L_p$  shifts along the straight line connecting the two extreme positions of  $L_p$ . This property makes it relatively easy to construct fig. 5 by interpolation.

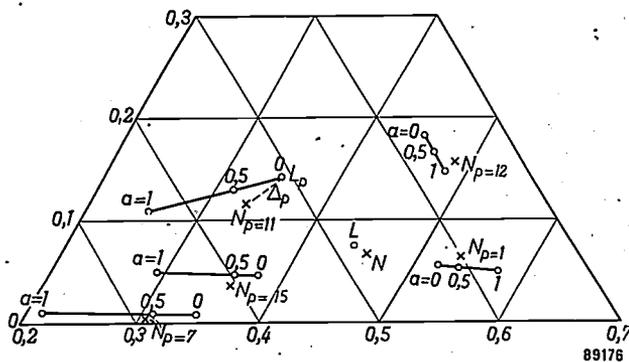


Fig. 4. Part of the Uniform Chromaticity Scale diagram after Judd<sup>10)</sup>. This diagram — which is derived by linear transformation from the chromaticity diagram of fig. 1 (see for example the book by Bouma cited in<sup>2)</sup>) — is such that equal distances correspond to equal differences in colour sensation (equal number of steps in which the colour difference is just perceptible).

The points  $N_p$ , with  $p = 1-7-11-12-15$  are the colour points of five test objects (coloured objects  $p$ , see fig. 5), illuminated by a black-body radiator at 4200 °K. At the same time, the colour points  $L_p$  are given for each object, these representing the colour it assumes when illuminated by means of a white fluorescent lamp with magnesium arsenate added to the phosphor, for various values of  $a$ , the proportion in which the two red phosphors are mixed. The colour points obtained by varying  $a$  always lie on straight lines.  $N$  is the colour point of the black body mentioned,  $L$  is that of the fluorescent lamp.

be seen from fig. 5 that for the lamp colour White, there is a region in the neighbourhood of  $a = 0.5$  where most of the colour discrepancies  $\Delta_p$  are markedly smaller than for higher or lower values of  $a$ .

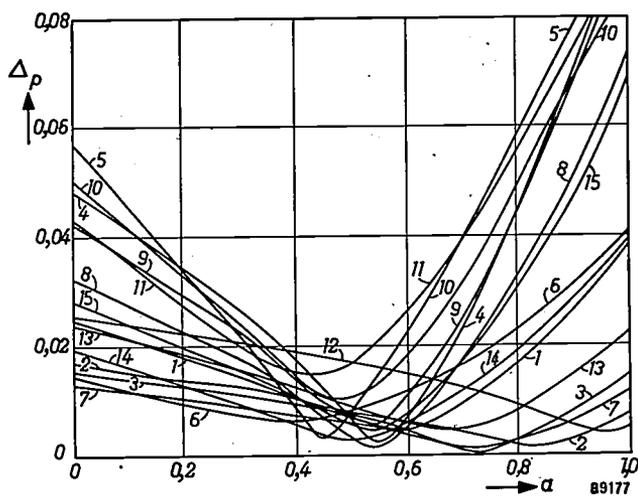


Fig. 5. Colour discrepancies  $\Delta_p$ , represented as distances in the UCS triangle, for coloured objects with the numbers  $p = 1$  to 15, illuminated by a black-body radiator at 4200 °K and then by a white fluorescent lamp with the mixing parameter  $a$ . For a specific value of  $a$ ; in this case  $a \approx 0.5$ , the colour differences are at a minimum. The greatest remaining colour discrepancy  $\Delta_m$  — in the case in point  $\Delta_m \approx 0.017$  — indicates the best colour rendering that can be attained by adding magnesium arsenate.

The objects  $p = 1$  to 6 are the pigments green, blue, yellow, orange, red and purple, whose spectral reflection curves are given in Hardy's Handbook of Colorimetry (M.I.T., Cambridge, Mass., 1936);  $p = 7$  to 15 are the colours pa 04, pa 17, pa 25, pa 33, nc 42, nc 54, nc 71, pa 88 and ia 21 from Ostwald's Colour Atlas.

The greatest remaining colour discrepancy for  $a = 0.5$  is  $\Delta_m = 0.017$ .  $a$  will thus be chosen to lie in this region, and the other mixing-proportions of the phosphor mixture then follow from the desired colour point, which is selected to have the same value as that of the standard White lamp.

In narrowing down the choice of  $a$ , the desirability of having a high efficiency must be borne in mind. True, the efficiency is only a secondary consideration here, but it is one nevertheless which cannot be totally ignored: a relatively small reduction in  $\Delta_m$  is not justified if it would mean a great sacrifice in luminous efficiency. Thus the White de Luxe lamp was developed with the phosphors mentioned and  $a = 0.34$ . The greatest colour discrepancy occurring with this lamp in the UCS triangle is  $\Delta_m = 0.023$ , the Harrison figure of merit is 76.

It should be noted that a little magnesium arsenate is also added in the White standard-type lamp ( $a \approx 0.1$ ). This gives rise to the low peak in the red in the spectral distribution of this lamp, see fig. 3c.

The colour discrepancy  $\Delta_m$  of the lamp colour "Warm White" can be reduced in an analogous manner to 0.02 for a value of  $a = 0.5$ . With the Daylight type lamp the colour discrepancy can be reduced to as low as 0.0075, for  $a \approx 0.5$ , but in so doing the question of efficiency again becomes significant, so that it has not been found desirable to go further than  $a \approx 0.30$ , with  $\Delta_m \approx 0.025$  for the so-called "Colour Matching" lamp (the colour point being slightly different from that of the standard-type lamp, viz. colour temperature 6500 °K). The values of  $\Delta_m$  and the corresponding figures of merit are collected in Table III.

Table III. Largest remaining colour discrepancies  $\Delta_m$  and the figures of merit  $H$  for the three types of "de Luxe" lamps which were obtained by improving the spectrum in the red with the aid of magnesium arsenate;  $a$  = ratio between the two red phosphors.

	$a$	Improved		Standard	
		$\Delta_m$	$H$	$\Delta_m$	$H$
"Cool Daylight"	0.3	0.025	75	0.055	67
"Warm White"	0.5	0.020	84	0.065	62
"White"	0.34	0.023	76	0.05	64

The improved "Warm White" lamp has never been realized in this form; with  $a = 0.5$  the efficiency would be undesirably low. (The calculation for this lamp, was made starting from a colour temperature of 3100 °K, colour point  $x = 0.427$ ,  $y = 0.396$ , which was laid down earlier.) The "Colour Matching" lamp has a colour point corresponding to a colour temperature of 6500 °K.

It will be observed that the addition of the red magnesium-arsenate phosphor — in roughly optimum proportions — to all three types of lamps leads to appreciable improvement of the colour rendering

qualities. Nevertheless the resulting lamps still do not satisfy the highest demands that can be made in this respect. Attempts were made to approximate more closely to the ideal by taking additional measures to correct the second deficiency mentioned above — the excess in the blue-violet and in the yellow-green.

#### *The absorption of the blue-violet mercury line*

An obvious means to reduce the blue-violet hump is to cause a sufficiently strong absorption of the mercury line in the lamp itself. This must be accompanied by an appropriate alteration of the phosphor mixture to eliminate the hump in the yellow-green spectral region, since this hump was only necessary in order to counter the displacing effect of the mercury line on the colour point. The notorious false rendering of yellow and orange objects will then no longer be found.

Experiments involving absorption of the mercury line were being made in Eindhoven as long ago as 1938-39. At that time G. Zecher prepared experimental lamps with uranium-glass walls which absorbed in the desired spectral region. The feasibility of the principle was hereby clearly demonstrated, but its use in practice came up against the impossibility of making tubes with a uniform glass wall of predetermined thickness to regulate the absorption. Experiments on absorbing lacquer layers applied to the outside of the lamp and containing suitable organic dyestuffs failed owing to the same difficulty and owing to well-nigh unavoidable local variations in the lacquer layer. The most attractive solution that remained was to use an inorganic pigment. Many such pigments do not need to be processed in a lacquer; they can be applied to the inner side of the glass wall, since they will withstand the heat treatment during the manufacture of the lamps without decomposing. Here, however, it proved difficult to find pigments with the desired spectral absorption curve.

The investigation took a turn for the better when it was observed that magnesium arsenate, which had been used in supplementing the fluorescence in the red, could be excited not only by the ultra-violet mercury radiation but also by the blue-violet spectral line. This meant that the undesirable radiation was absorbed and that useful radiation was obtained for it. As normally prepared, magnesium arsenate phosphor absorbs the blue-violet line only very weakly but by adopting a special mode of preparation, this absorption can be increased to such an extent that a layer of roughly the same thickness as the normal fluorescent layer

(a few milligrams per cm<sup>2</sup>) is sufficient to bring about the desired absorption (fig. 6). The use of two layers of the phosphor, one absorbing and one non-absorbing, makes it possible to regulate to a

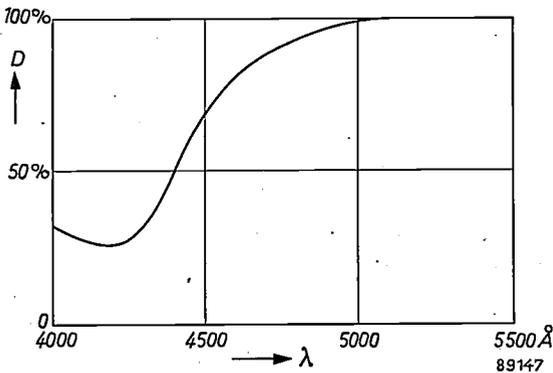


Fig. 6. Spectral absorption curve of magnesium arsenate. The transmission  $D$  is plotted as a function of the wavelength  $\lambda$ . The absorption factor referred to in the text is  $\delta = |\log D|$ . For non-scattering substances this factor is proportional to the thickness of the layer. This likewise applies approximately to magnesium arsenate, although here the absorption is attended with much scattering and with reflection.

certain extent the relative absorption of blue-violet and the fluorescent contribution in the red, independently of each other.

The question now arose as to whether the colour rendering would improve indefinitely with increasing absorption or whether there was an optimum absorption in this respect. Calculations showed that the latter was the case. The calculations were carried out in a manner analogous to that described above in regard to the addition of red; since they are rather time-consuming, we have confined ourselves to one phosphor mixture, the mixture of halophosphate, magnesium arsenate and willemite already referred to. We shall give here a brief outline of these calculations.

A value for the absorption is first chosen. This is characterized by the value  $\delta = |\log D|$ , where  $D$  is the transmission factor for the mercury line  $\lambda = 4358 \text{ \AA}$ . The transmission curve of the absorbing layer (transmission factor as a function of the wavelength) for the chosen value of  $\delta$  is derived from fig. 6. The ordinates of the emission curves for the four phosphors used (see above) for each wavelength are then multiplied by the transmission factor, so that four new, effective emission curves are found. Starting from chosen values for the mixing proportion  $a$  of the two reddest phosphors, the colour discrepancies  $\Delta_p$  of the 15 test objects are calculated as before as a function of  $a$  (cf. figs. 4 and 5). It is then possible to establish which value for  $a$  is the best, and how great is the maximum remaining colour discrepancy  $\Delta_m$  associated with that value.

This determination of the best possible colour rendering was repeated for various values of  $\delta$ , and this whole series of calculations was carried out for mixtures of the ("effective") phosphors corresponding to the three lamp colours "Cool Daylight", "White" and "Warm White". The bulky numerical material amassed was processed to give three graphs fig. 7a, b and c. In these graphs  $a$  is plotted along the abscissa and  $\delta$  along the ordinate, and lines are drawn for constant values of the maximum colour discrepancy  $\Delta_m$ .

To interpret the graphs we look first at the line  $\delta = 0$  (abscissae), which corresponds to the case treated earlier with no absorption. In fig. 7a which refers to the Cool Daylight lamp, we see that for  $a = 0.35$ , the value of  $\Delta_m$  is 0.02; for  $a = 0.49$ ,  $\Delta_m = 0.0075$ , while for still higher values of  $a$ ,  $\Delta_m$  increases, e.g.  $\Delta_m = 0.02$  for  $a = 0.6$ . If we then take cases in which  $\delta > 0$  (with absorption), it can be seen in fig. 7a that for  $\delta = 0.05$ , the discrepancy  $\Delta_m$  is likewise reduced to 0.0075,  $a$  now being 0.47, but that  $\Delta_m$  does not assume still

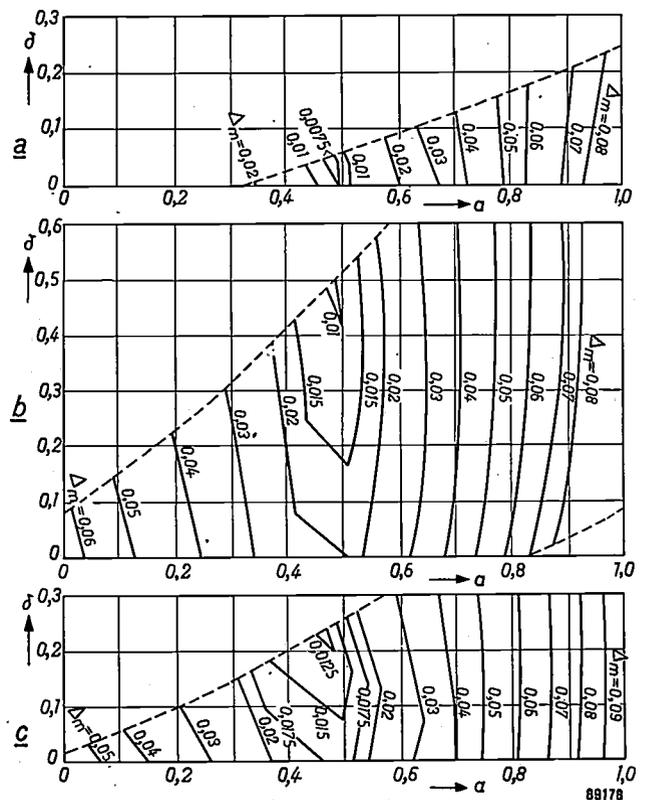


Fig. 7. The greatest remaining colour discrepancy  $\Delta_m$  in the UCS diagram calculated for the lamp colours a) Cool Daylight, b) Warm White, c) White, with the phosphors halophosphate, magnesium arsenate and willemite; the mixing parameter  $a$  is varied and moreover a partially absorbing layer of magnesium arsenate, absorption factor  $\delta$  for  $\lambda = 4358 \text{ \AA}$ , has been applied in order to attenuate the blue-violet mercury line. Lines of constant  $\Delta_m$  in the  $a$ - $\delta$  plane are drawn. Outside the broken boundary lines the light contribution of one of the phosphors has negative values, so that the prescribed colour points cannot be realized with those combinations of  $a$  and  $\delta$ .

lower values. This is in agreement with what was to be expected: it has already been pointed out that with the Cool Daylight lamp the second deficiency (blue-violet and green-yellow humps) could not be expected to have any great effect. For this lamp, therefore, correction of this deficiency by absorption does not lead to any noteworthy improvement.

For the lamp colour Warm White the situation is quite different; see fig. 7b. Without absorption,  $\delta = 0$ ,  $\Delta_m$  can at best be reduced to 0.02, for a value of  $a \approx 0.5$ . With absorption, on the other hand,  $\Delta_m$  is reduced in the most favourable case to 0.01, i.e. just to the observable limit; the required value of  $\delta$  for this is 0.4 to 0.5, that of  $a$  is again approximately 0.5. Since this is one of the most important results of the whole investigation from a practical point of view, we have collected the numerical values of a number of points in the graph in fig. 7b in Table IV, together with Harrison's figure of merit for each phosphor mixture with given  $\delta$  and  $a$ . It will be observed that  $H$  figures of more than 90 can be attained<sup>13)</sup>.

Finally a few words regarding the lamp colour White, for which fig. 7c obtains. For  $\delta = 0$  (see also fig. 5) the smallest value of  $\Delta_m$  which can be obtained is 0.017, with  $a = 0.5$ . Optimum absorption appears to be  $\delta = 0.23$ , the maximum colour discrepancy with  $a = 0.46$  then being reduced to 0.012.

<sup>13)</sup> The fact that both Harrison's figures of merit and the residual maximum colour discrepancies  $\Delta_m$  are known for a large number of different types of light, makes it possible to test the validity of the figure of merit. With this in view the figure of merit has been plotted as a function of  $\Delta_m$  in fig. 8. Although a fairly large spread can be observed, there is nevertheless a clear correlation between the two quantities (the straight line shown). It can thus be seen that we are justified in employing Harrison's figure of merit, at least in the first instance, for assessing the colour rendering qualities of a fluorescent lamp.

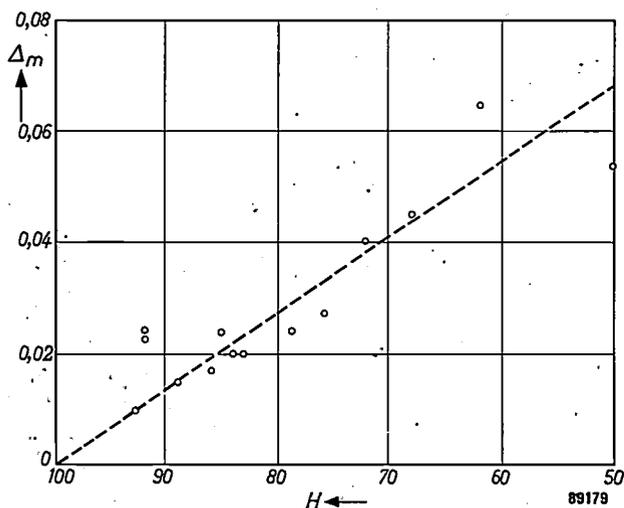


Fig. 8. Maximum colour discrepancy  $\Delta_m$  plotted against Harrison's figure of merit  $H$ , both calculated for a number of phosphor mixtures with the lamp colour Warm White; see Table IV.

Table IV. Calculated maximum colour discrepancies  $\Delta_m$  and figures of merit  $H$  for "Warm White" fluorescent lamps with halophosphate, magnesium arsenate and willemite, with various mixing parameters  $a$  for the red light contributions and a varying degree of absorption  $\delta$  of the blue-violet mercury line. (The blank spaces indicate combinations of  $a$  and  $\delta$  for which one of the constituent phosphors would have to give a negative contribution for the "Warm White" colour point, so that this colour point cannot be realized with these combinations.)

$\delta$	0		0.2		0.4		0.6	
	$\Delta_m$	$H$	$\Delta_m$	$H$	$\Delta_m$	$H$	$\Delta_m$	$H$
0	0.065	62	—	—	—	—	—	—
0.2	0.045	68	0.040	72	—	—	—	—
0.4	0.024	79	0.020	83	0.017	86	—	—
0.5	0.020	84	0.015	89	0.010	93	—	—
0.6	0.027	76	0.024	85	0.023	92	0.025	92
0.8	0.063	33	0.057	45	0.054	50	0.054	50

### Construction of the fluorescent de Luxe lamps; new phosphors

The following sub-sections summarize, with some amplifications, the results thus far achieved and explain their application in a number of de Luxe fluorescent lamps now on the market.

#### Daylight de Luxe or Colour Matching

The present-day Daylight de Luxe (or Colour Matching) lamp makes use of halophosphate with magnesium arsenate ( $a = 0.3$ ) and willemite, but there is no absorption. Its colour temperature is 6500 °K and its spectral energy distribution is as shown in fig. 9a. The maximum colour discrepancy is 0.025 and its luminous efficiency is 50 lm/W.

Although the lamp is satisfactory in practice, better colour rendering, if this could be attained without seriously detracting from the luminous efficiency, would certainly be desirable.

Since absorption is of little avail in this case, improvement must be sought in the application of other phosphors. The most obvious shortcoming that calls for correction is the fairly deep "valley" which still exists between the red maximum of the halophosphates (5800 Å) and that of the magnesium arsenate (6600 Å). A combination of different phosphors with cadmium borate as its principal component was found to answer this purpose well, but was unusable in practice, since the light yield of the borate falls off too rapidly during life.

#### Warm White de Luxe

In the Warm White de Luxe lamp the combination of halophosphate and magnesium arsenate is used, colour rendering being improved by a layer of magnesium arsenate which is applied between the fluorescent layer proper and the wall of the tube

and which reduces the intensity of the blue-violet mercury line to about  $\frac{1}{3}$  ( $\delta = 0.5$ ). The spectral distribution of this lamp is shown in fig. 9b; the colour temperature is 2900 °K, the maximum colour discrepancy 0.01, the figure of merit 92, the luminous efficiency 47 lm/W. Table V shows the spec-

trum divided up into spectral bands and includes for comparison the spectral distributions of a black body at 2900 °K and of an experimental Warm White de Luxe lamp without absorption. It will be observed that absorption has resulted in a much better approximation to the ideal in the blue-green

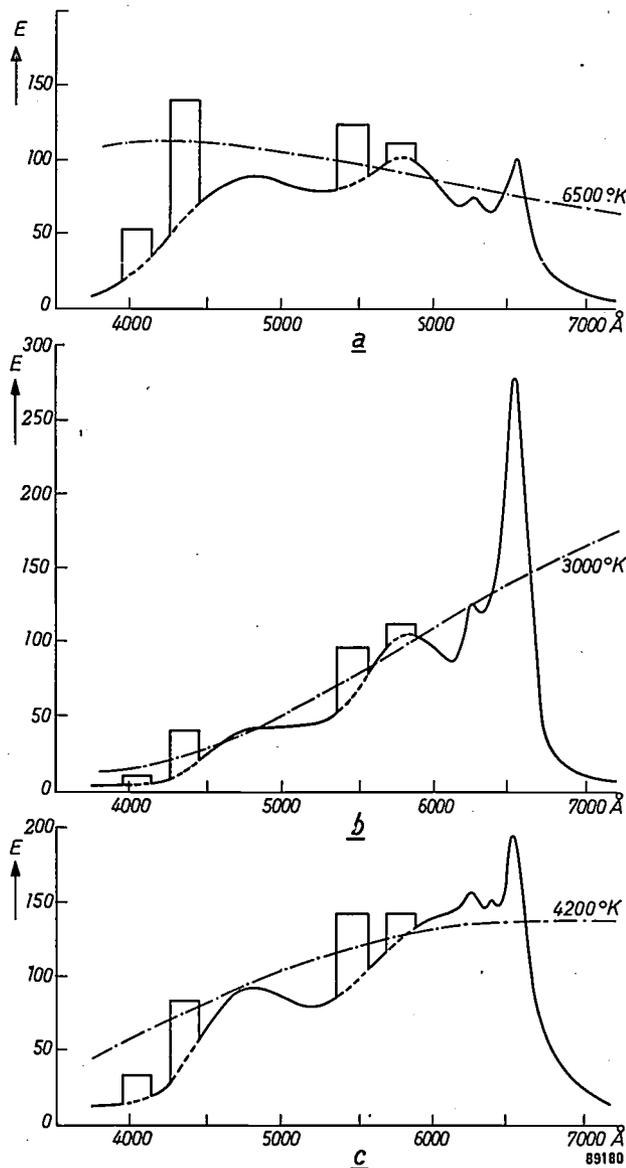


Fig. 9. Spectral energy distribution of fluorescent de Luxe lamps with good colour rendering, which are being produced at the present time.  
 a) Daylight de Luxe or Colour Matching, with halophosphate, magnesium arsenate and willemite. Colour temperature 6500 °K.  $H = 75$ ; luminous efficiency 50 lm/W.  
 b) Warm White de Luxe, with the same phosphors and an absorbing layer of magnesium arsenate ( $\delta = 0.5$  for  $\lambda = 4358 \text{ \AA}$ ). Colour temperature 2900 °K.  $H = 92$ ; luminous efficiency 47 lm/W.  
 c) White de Luxe, with calcium strontium silicate, blue halophosphate and magnesium arsenate together with an absorbing layer of magnesium arsenate ( $\delta = 0.3$ ). Colour temperature about 4000 °K.  $H = 95$ ; luminous efficiency 44 lm/W.  
 As in fig. 3a-c, the spectral energy distribution of the corresponding black body is indicated (dot-dash line).

Table V. Distribution of light-flux over eight spectral bands for fluorescent lamps "White de Luxe" and "Warm White de Luxe", the latter both with and without absorption of the blue-violet mercury line. All the lamps employ halophosphate, magnesium arsenate and willemite. For comparison the distribution for a black body with the corresponding temperature is shown.

Spectral band No.	Wave-lengths in Å	"White de Luxe" with absorption	4200 °K	"Warm White de Luxe" without absorption	2900 °K	"Warm White de Luxe" with absorption
		I	3800-4200	0.005	0.017	0.006
II	4200-4400	0.21	0.14	0.31	0.062	0.16
III	4400-4600	0.43	0.50	0.08	0.26	0.22
IV	4600-5100	8.1	8.1	2.0	5.5	4.5
V	5100-5600	37.6	38.6	36.1	33.7	34.1
VI	5600-6100	39.5	39.1	45.5	42.5	44.7
VII	6100-6600	13.7	12.5	15.6	16.4	15.5
VIII	6600-7800	0.57	0.99	0.48	1.54	0.87
Figure of merit $H$		95	100	81	100	92

spectral bands (III and IV). At the same time it will be seen that it has been possible to enhance the red content (spectral bands VII and VIII) without incurring an increase in the green (V and VI) to keep the colour point constant.

This lamp has been found in practice to come up to all expectations both for use in restaurants, department stores etc., and for use in the home.

White de Luxe

With the mixture of halophosphate and magnesium arsenate it has been found possible to produce a fairly good White de Luxe lamp, especially if absorption is incorporated (up to  $\delta = 0.23$ ). However, it has been found possible to obtain particularly good results by using a phosphor mixture containing a red silicate phosphor, and halophosphates in combination with a layer of absorbing magnesium arsenate which cuts down the intensity of the blue-violet mercury line to about  $\frac{1}{2}$ . The lamp made with this mixture (colour temperature about 4000 °K) is found to give even better colour rendering than the new Warm White de Luxe lamps. Its spectral distribution is shown in fig. 9c and the distribution of light flux over the eight spectral bands is given in Table V. The figure of merit is 95 and the luminous efficiency 44 lm/W.

The outline of the development of fluorescent lamps drawn above is not the whole story; it makes no mention of various sidetracks which were followed up for a short while in the laboratory (and occasionally in manufacture) before it was realized that they were leading to a dead end. This account will perhaps have given too strong an impression that in the development of these lamps, the course to be followed was plotted solely by the compass of calculation. If this is so a slight rectification is called for. By making the concept of "colour" with its many physiological complications accessible to calculation, Bouma and other workers in this field have certainly brought us a big step forward; but subjective approval by the user remains the ultimate goal of good colour rendering by any type of lighting. Wide-scale testing under practical conditions, therefore, is the deciding factor in determining whether any new type of lamp shall be manufactured.

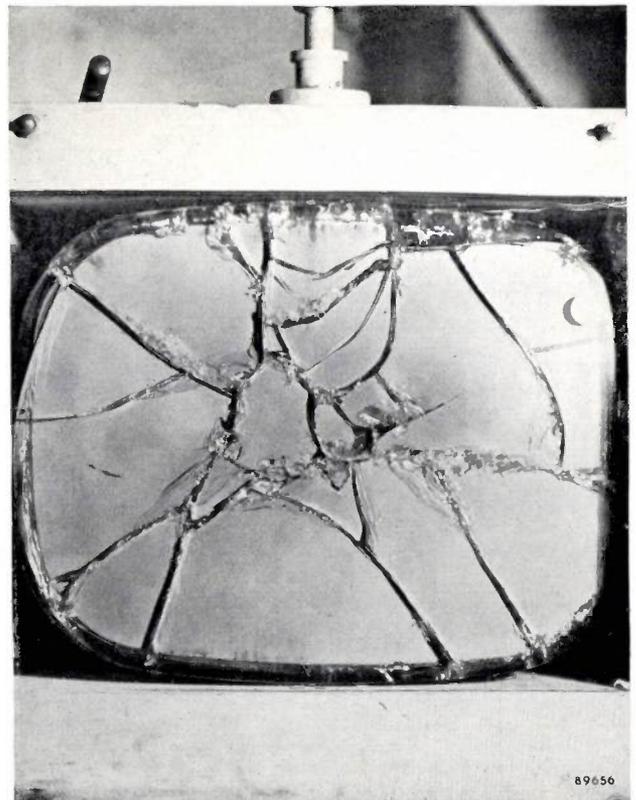
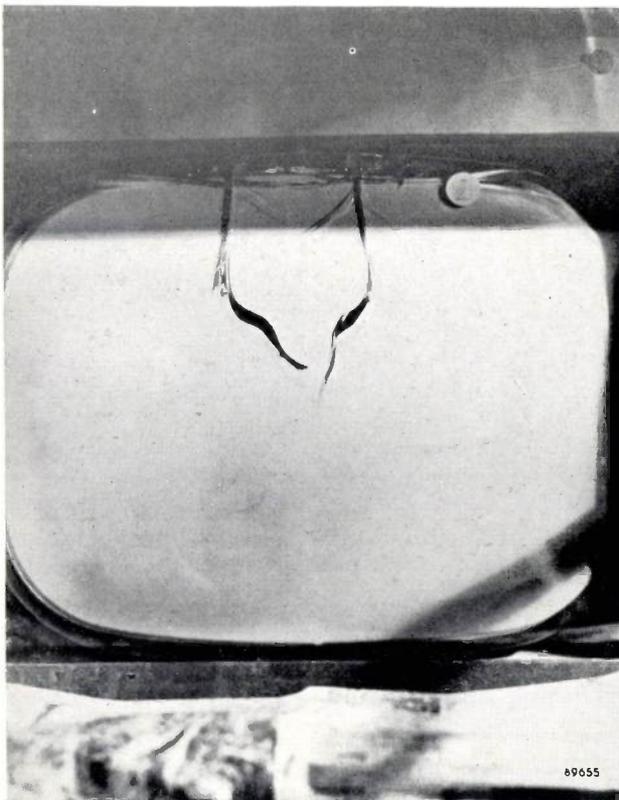
**Summary.** By using a mixture of various phosphors in a tubular fluorescent lamp, the colour point of the lamp in the chromaticity diagram can be varied within wide limits. Three colour points have been chosen for lamps which correspond either to average daylight (colour Cool Daylight, colour temperature 6800 °K), or to incandescent lamp light (colour Warm White, 2900 °K), or to light which lies between the two (colour White, 4200 °K). In addition to the three standard types of fluorescent lamps corresponding to the above, which employ halophosphate phosphor and have optimum luminous efficiencies, other types have been developed, the "de Luxe" lamps (with approximately the same colour temperatures), which have somewhat lower luminous efficiencies, but whose colour rendering has been made as natural as possible. The colour rendering qualities of these lamps have been improved by adding red to the spectrum and attenuating the blue-violet mercury line (this being attended by a reduction of the light contribution in the yellow green). Both effects can be achieved with the aid of a magnesium arsenate phosphor which is prepared in different ways. The general form of the calculations for the determination of the best proportions of mixing and optimum degrees of absorption is outlined. Very good results have been arrived at in this way for the colours Warm White de Luxe and White de Luxe. As regards the Daylight de Luxe or "Colour Matching" lamp, the fact that the luminous efficiency cannot be allowed to suffer unduly for the sake of improvements in the colour rendering makes it preferable to find a solution through a different mixture of phosphors. A new mixture of phosphors combined with absorption by magnesium arsenate has given good results in a new White de Luxe fluorescent lamp.

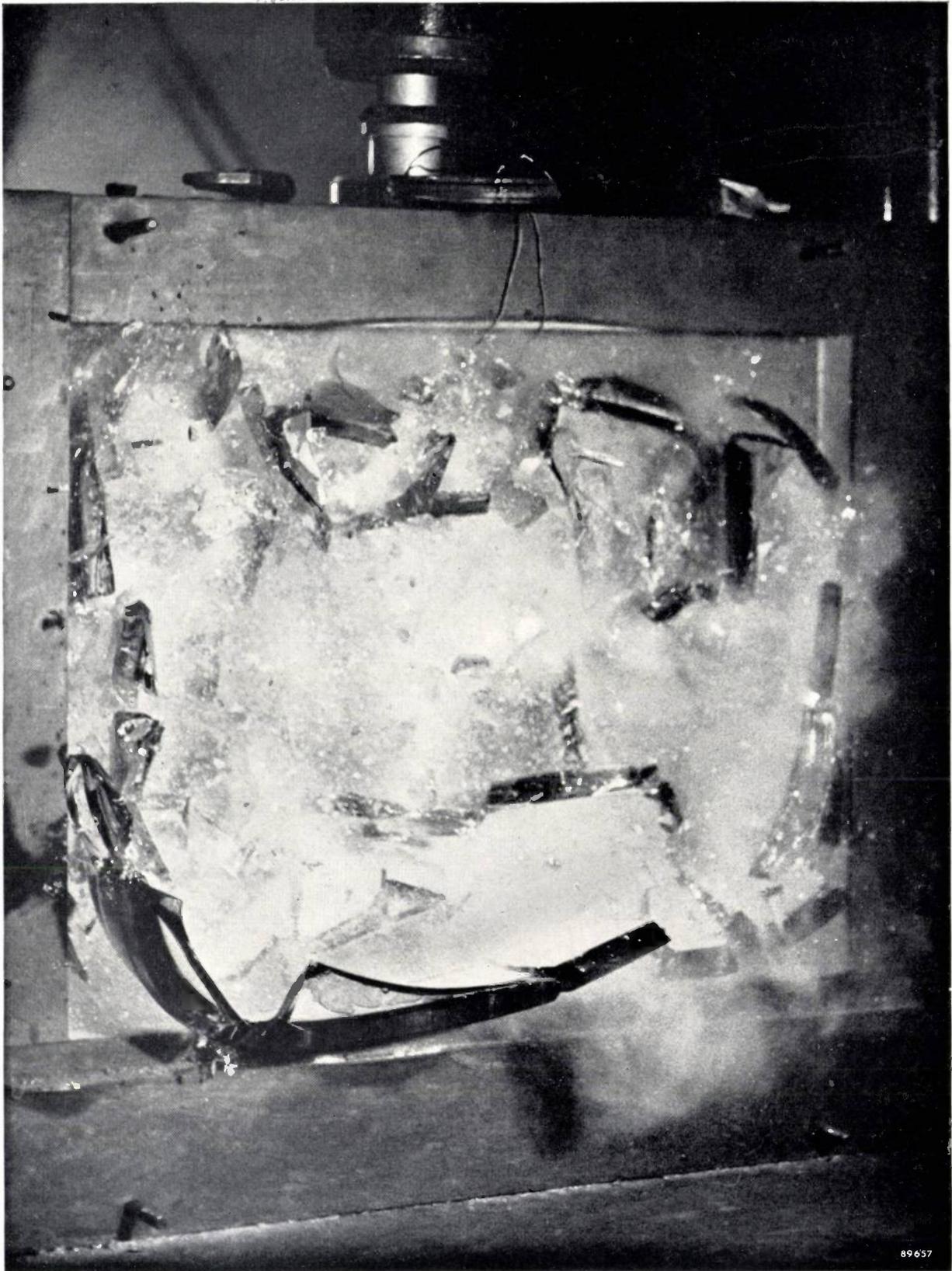
### HIGH-SPEED PHOTOGRAPHS OF THE IMPLOSION OF TELEVISION PICTURE TUBES

The photographs shown here are three of a series taken in connection with tests on the implosion of television picture tubes. The photographs were taken by R. A. Chippendale at the Mullard

Research Laboratories, Salfords, Surrey, England.

Implosion was initiated by a steel weight dropped on to the rim of the tube. A conducting graphite strip on the glass triggered the flash-tube (Mullard





LSD 2) at an early stage of fracture, immediately after impact. Although these photographs are not of the *same* event they broadly represent three steps in the implosion: 1) Initial stage, immediately

after impact; 2) 10 msec delay: screen shattered; 3) 100 msec delay: tube disintegrates. In this last case, the usual implosion guard (a glass or acrylic sheet fixed in front of the tube) was removed.

## COUNTERS FOR X-RAY ANALYSIS

by P. H. DOWLING \*), C. F. HENDEE \*), T. R. KOHLER \*) and W. PARRISH \*).

621.387.4:545.824:548.734

*Geiger counters, proportional counters and scintillation counters have played an all-important role as radiation detectors in nuclear physics and allied fields for many years. The Geiger counter was the first of these to enter the field of soft X-rays, and in recent years the other types of counters have been introduced. Faced with a choice between the three types, the worker in the field must know their relative merits and limitations for different applications. Such an assessment is given in this article, on the basis of the physical properties of the counters and of practical measurements under a variety of conditions \*\*).*

### Introduction

About 15 years ago Geiger counters adapted to the detection of relatively soft X-rays were introduced into the general X-ray diffraction field as a substitute for photographic film. The most striking feature of the counter tube in this field is its superiority to film in the evaluation of relative radiation intensities. This has led to the successful development of the method of "diffractometry" and related procedures. Three articles on this subject recently published in this Review may be quoted <sup>1) 2) 3)</sup>.

The development of the Geiger counter for X-ray detection has been followed in recent years by the adaptation of the *proportional counter* and the *scintillation counter* to the same purpose <sup>4)</sup>. Detectors of this type and their associated circuits, suitable for X-ray diffractometry and X-ray spectrochemical analysis, have been developed and engineered in the Philips Laboratories at Irvington in such a form that they can be readily substituted for the Geiger counter tube on the "Norelco" X-ray goniometer. Each of the three detectors thus available has its specific advantages and, of course,

limitations. In *fig. 1* the three detectors and the complete interchangeable units are shown. A description of the new detectors will be given and their underlying mechanisms and essential characteristics will be discussed in this article.

### The Geiger counter

For the sake of comparison, the design of a Geiger counter tube for X-ray detection, which has been described on several occasions in this Review, should be briefly recapitulated.

A cross-section of the standard "Norelco" Geiger counter tube, type 62019, is shown in *fig. 2*. The tube consists of a chrome-iron cathode cylinder 2 cm in diameter, along whose axis a 0.75 mm tungsten anode wire is mounted. The tube is filled with argon gas to a pressure of 55 cm Hg, and a small amount (less than 0.4%) of chlorine is added as a quenching agent. Approximately 1400 V is applied between cathode and anode. X-rays enter the tube axially through an end window of mica 0.013 mm thick. An X-ray quantum will give rise to a discharge pulse in the tube if it is absorbed in the gas within the active volume, which is a cylinder around the anode wire about 10 cm long (practically the whole length of the tube) and 1.2 cm in diameter. When copper  $K\alpha$  radiation ( $\lambda = 1.54 \text{ \AA}$ , quanta of energy 8 keV) is used, in a beam not wider than the active volume, about 60% of the X-ray quanta impinging on the tube window are detected. This percentage is called the *quantum counting efficiency* of the counter tube.

The quantum counting efficiency depends on the absorption of the radiation by the active gas column. The effect of absorption by the tube window and by the adjoining inactive part of the gas column (which, however, in the 62019 type of tube is

\*) Philips Laboratories, Irvington-on-Hudson, N.Y., U.S.A.

\*\* The subject-matter of this article was presented in part at the 3rd Congress of the International Union of Crystallography, Paris, July 1954. Cf. also *Acta Cryst.* 7, 626, 1954.

<sup>1)</sup> W. Parrish, E. A. Hamacher and K. Lowitzsch, The "Norelco" X-ray diffractometer, *Philips tech. Rev.* 16, 123-133, 1954/55.

<sup>2)</sup> W. Parrish, X-ray intensity measurements with counter tubes, *Philips tech. Rev.* 17, 206-221, 1955/56 (No. 7-8).

<sup>3)</sup> W. Parrish, X-ray spectrochemical analysis, *Philips tech. Rev.* 17, 269-286, 1955/56 (No. 10).

<sup>4)</sup> Proportional counter: S. C. Curran, J. Angus and A. L. Cockroft, *Phil. Mag.* 40, 36-52, 1949.

G. C. Hanna, D. H. W. Kirkwood and B. Pontecorvo, *Phys. Rev.* 75, 985-986, 1949.

W. Bernstein, H. G. Brewer Jr. and W. Rubinson, *Nucleonics* 6, Feb. 1950, pp. 39-45.

Scintillation counter: F. H. Marshall, J. W. Coltman and A. I. Bennett, *Rev. sci. Instr.* 19, 744-770, 1948.

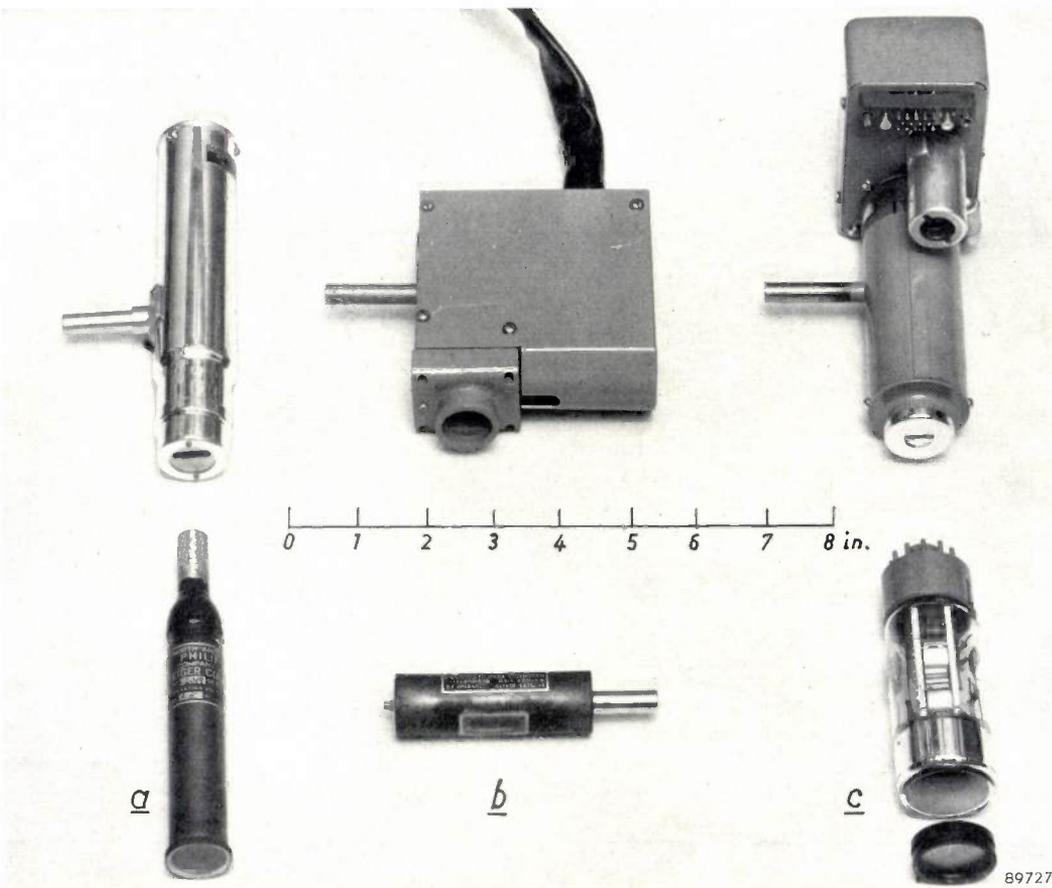


Fig. 1. Interchangeable X-ray detectors for the "Norelco" equipment. a) Geiger counter, type 62019. b) Xenon-filled proportional counter, type 62031. c) Scintillation counter, type 52245.

The proportional and scintillation counter have cathode follower circuits, which are mounted directly with the counter on the goniometer arm. The complete detector units are shown in the upper part of the photograph.

only about 3 mm long) limits the efficiency. Since both the desirable and the undesirable absorption are a function of the wavelength of the radiation, the quantum counting efficiency depends strongly on the wavelength. This is a most important factor for practical applications, as will be discussed below.

The spectral response of the argon-filled Geiger counter tube, type 62019, as calculated from absorption data<sup>5)</sup> and which has been confirmed by measurements<sup>6)</sup>, is shown in *fig. 3*.

The detailed mechanism of quantum detection in a Geiger counter tube can be described briefly as follows<sup>7)</sup>. Absorption of a quantum by a gas atom

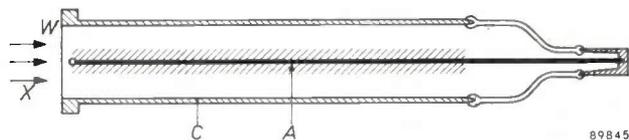


Fig. 2. Cross-section of "Norelco" Geiger counter tube, type 62019. The tube is filled with argon (55 cm Hg) and a small amount of chlorine. A voltage of about 1400 V is applied between cathode C and anode A, the permissible voltage range ("plateau") is more than 300 V wide. X-rays (X) enter axially through an end window W, the active volume of the tube is indicated by shading. The useful life of the tube is practically unlimited; even after having counted 10<sup>9</sup> quanta, a tube of this type showed no deterioration of characteristics.

<sup>5)</sup> J. Taylor and W. Parrish, *Rev. sci. Instr.* **26**, 367-373, 1955 (No. 4).

<sup>6)</sup> W. Parrish and T. R. Kohler, *Rev. sci. Instr.* **27**, 795-808, 1956 (No. 10).

<sup>7)</sup> See for instance: D. H. Wilkinson, *Ionization chambers and counters*, Cambridge Univ. Press, 1950. Also: N. Warmoltz, *Philips tech. Rev.* **13**, 282-292, 1951/52. — It should be noted that Geiger counter detectors for nuclear experiments in which  $\beta$ -rays or very hard X-rays ( $\gamma$ -ray quanta of millions of eV energy) are to be measured, have a somewhat different mechanism: the quanta in these cases are absorbed — generally to a rather low percentage — in the cathode wall or in the anode, not in the gas. The development of the discharge is similar to that in the detector for soft X-rays described here.

results in the ejection of a photo-electron. The gas ion so produced gets rid of its excess energy in most cases by emitting one or more additional high-energy electrons (internal conversion by the Auger process).

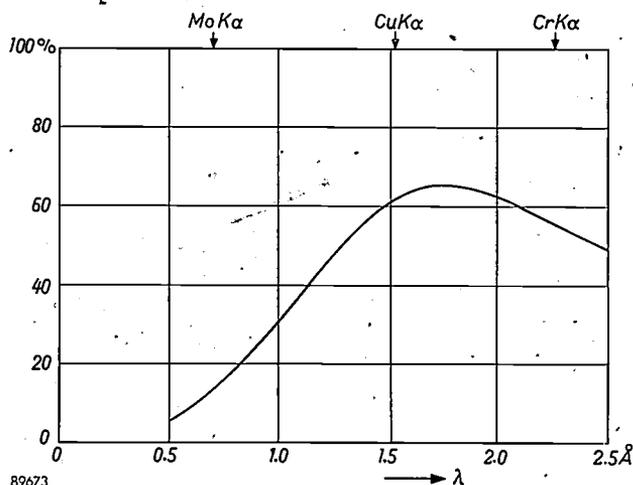


Fig. 3. Quantum counting efficiency as a function of wavelength (spectral response) for the argon-filled Geiger counter, calculated from absorption data<sup>8)</sup>.

These conversion electrons as well as the photo-electron in turn spend their kinetic energy by forming ion pairs in the gas. Thus, a total number of about 310 ion pairs are formed in argon by the absorption of one copper  $K\alpha$  quantum. Each of the secondary electrons will give rise to a Townsend avalanche owing to acceleration in the electric field of the counter tube. Metastable atoms and photons emitted from the avalanches cause the emission of still more electrons throughout the tube, so that a self-sustained discharge filling the whole volume of the tube is started. The discharge is quenched after a short time either by the action of an external quenching circuit causing the tube voltage to drop as the discharge current is built up, or by the action of the quenching agent added to the gas (self-quenching).

The latter action can be explained as follows. Most of the ionization occurs near the central wire, where the electric field strength is largest. The electrons thus formed are quickly attracted to (collected at) the central wire, leaving behind a positive ion space-charge sheath around the wire, which effectively lowers the electric field strength in the gas and momentarily prevents further ionization. The positive ion sheath drifts out from the wire, restoring after a short time ("dead time") the original field conditions. This would allow the discharge to be reinitiated by other electrons subsequently emitted from the wall; such an emission, however, is prevented by the quenching agent.

The gas amplification process described results in discharge pulses of large amplitude, say 1-10 V, a feature which makes the Geiger counter an extremely useful radiation detecting device for many problems. The "dead time", on the other hand, limits the usefulness of this device. During this time another quantum entering the tube cannot give rise to a separate pulse and so is not detected. As a result a proportion of the arriving quanta is not counted, and this proportion increases with their arrival rate, i.e. the response becomes non-linear at large radiation intensities. The dead time of the Geiger counter type 62019 is 170  $\mu$ sec; the "effective" dead time for  $CuK\alpha$ -radiation produced by an X-ray tube operated on a 35 kV (peak) full-wave rectified voltage<sup>8)</sup> is 270  $\mu$ sec, so that appreciable deviations from linearity ( $> 8\%$ ) occur at counting rates as low as 500 counts per second. Individual intensity values (peak values of diffraction lines) can be corrected for the non-linearity effect. No practical correction is possible, however, when the number of quanta is measured during a period of *varying* radiation intensity, as, for example, in the measurement of "integrated line intensities"<sup>9)</sup>.

#### The proportional counter

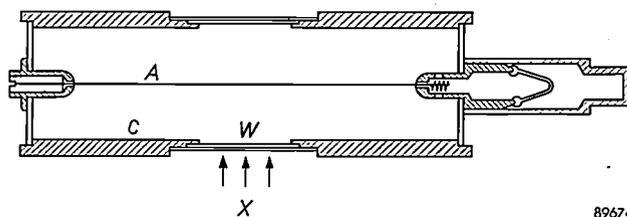
The large pulse size and the long inactive period of the Geiger counter both originate from the spreading of the discharge throughout the volume of the tube. In the proportional counter, whose design is basically similar to that of the Geiger counter, the voltage on the tube is chosen *below* the threshold value for the self-sustained discharge (i.e. the lowest limit of the Geiger plateau), and the discharge is then essentially confined to a plane normal to the axis of the tube, located where the original X-ray quantum was absorbed. The discharge current in this case is only the sum of the Townsend avalanches of the secondary electrons. The "gas amplification factor" therefore is about  $10^4$ , as against  $10^6$  or  $10^7$  in the Geiger counter tube, and the pulses produced are rather small. Because of the limited extent of the discharge, however, nearly the whole volume of the tube remains active and can detect a second X-ray quantum absorbed shortly after the first. In fact, two quanta need be separated only by about 0.2  $\mu$ sec, i.e. the electron collection time, in order to be separately detected. When the average current of the counter tube is used as a measure of the quantum arrival rate, it is even unnecessary to have separated

<sup>8)</sup> Loc. cit.<sup>2)</sup>, p. 218.

<sup>9)</sup> Loc. cit.<sup>2)</sup>, p. 216.

pulses; linearity is then maintained up to extremely high counting rates, say  $10^7$  counts/sec or more. (At still higher rates, the increasing space charge lowers the effective field.)

Since the pulses produced in the proportional counter are approximately  $10^3$  times smaller than those of a Geiger counter, an amplifier having a gain of approximately  $10^3$  must be added in order to bring the pulses to the level required for operating the counting circuits (about 0.25 V). Evidently, this is the price paid for better linearity.



89674

Fig. 4. Cross section of "Norelco" proportional counter, type 62031. *A* anode wire, *C* cylindrical cathode. This counter is filled with xenon (32 cmHg); a small quantity of methane is added as a discharge-limiting agent. About 1700 V is applied, the required voltage being dependent on the radiation measured. X-rays *X* enter through a side window *W*. Unabsorbed quanta leave the tube through a window diametrically opposite the entrance window. (Not all proportional counters are fitted with this exit window.)

A well-known property of the proportional counter, which is responsible for its name, is the proportionality between the size of the pulse and the energy of the absorbed quantum. In fact, whereas the pulses of a Geiger counter vary only slightly in size, the original quanta acting merely as triggers, the pulses delivered by a proportional counter each represent a charge equal to a fixed multiple (the gas amplification factor) of the ionization produced by the absorbed quantum, and hence are proportional in size to the quantum energy. This proportionality is a very useful characteristic since it enables the elimination of undesired radiation components by the method of pulse height discrimination. We shall deal with this point at some length in this article.

The "Norelco" proportional counter pictured in fig. 1, and whose cross-section is shown in fig. 4, is fitted with a 0.013 mm mica window of size  $0.6 \times 1.6$  cm in the side wall of the cathode cylinder, with the X-ray beam entering perpendicularly to the wire. This design is adopted to ensure the proportionality just mentioned: in order to make the gas amplification factor for all quanta as nearly equal as possible, the spatial distribution of the electric field should be identical for all quanta, irrespective of the place where they are absorbed. (At the ends of the cylindrical tube considerable distortions of the

equipotential surfaces are unavoidable.) Moreover, the "Norelco" tubes are made approximately four times longer than the diameter, in order to enhance the uniformity of the field distribution in the vicinity of the window, and any disturbing effect of the mica window itself is avoided by covering its interior wall with either a very thin film of evaporated gold or a thin sheet of beryllium. In some tubes an exit window of size  $0.8 \times 1.7$  cm is placed diametrically opposite the entrance window to allow the unabsorbed portion of the X-ray beam to leave the tube without striking the chrome-iron cylindrical cathode wall and possibly causing X-ray fluorescence.

The unabsorbed portion of the beam of course should be made as small as possible. This was the very reason why in the early days of diffractometry the end window design was introduced for the argon-filled Geiger counter X-ray detector, ensuring a long active gas column. The shorter length of the absorption path in the side-window design is compensated by selecting a heavier gas, in most cases *xenon*. The proportional counter, type 62032, filled with pure xenon to a pressure of 32 cmHg, has practically the same quantum counting efficiency for  $\text{CuK}\alpha$  as the Geiger counter 62019. In fact, the efficiency of both tubes is nearly identical in the whole spectral region between about 0.5 and 2.5 Å; cf. fig. 5. Higher gas pressures are not desirable because the operating voltages increase rapidly with pressure and would exceed the range of the "Norelco" power supply.

If the internal gas amplification at which the counter is operated is increased too far, the tube may go into continuous discharge. Moreover, even before this point is reached, oscillations in the form

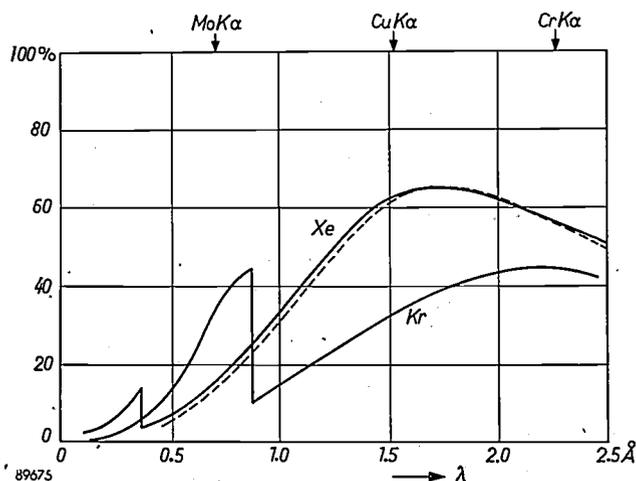


Fig. 5. Spectral response of the proportional counter filled with xenon (curve *Xe*) and with krypton (curve *Kr*), calculated from absorption data<sup>5</sup>. The dotted line is the spectral response of the Geiger counter from fig. 3.

of "after-pulses" occur after each normal X-ray pulse. The relative amplitudes of the after-pulses increase with voltage and may be sufficient to operate the scaling circuit, thereby falsifying the measured counting rate. It is therefore essential to keep the internal gas gain below some maximum value. It has been our experience that the operating voltage therefore should be chosen such as to produce pulses not exceeding 1 millivolt. (This should apply to the most energetic X-ray quanta — desired or not — which are absorbed in the counter, since the amplitude of the after-pulses increases with the quantum energy.) Hence it is necessary in practice to use a rather high external (amplifier) gain.

The after-pulsing phenomena and the difficulties involved in using high gain amplifiers (especially when pulse height discrimination is used) can be avoided by the addition of methane or some other organic gas to the noble gas of the proportional counter. The "quenching" or "limiting" agents prevent secondary processes and limit the discharge to a single avalanche. Thus, with the "quenched" proportional counter ("Norelco" type 62031) a higher gas amplification can be used, which allows much greater flexibility in the equipment and technique.

Since the limiting agent will gradually decompose, the quenched proportional counters have a limited useful life, whereas pure gas counters have an extremely long life. The useful life attained by tubes with not more than 10% methane, however, is adequate for all practical purposes (at least 1 or 2 years, in average practice).

Proportional counters filled with *krypton* have been used for the detection of  $\text{MoK}$  radiation. The spectral response curve for the krypton-filled counter, which is also shown in fig. 5, reveals a better quantum counting efficiency in the  $0.8 \text{ \AA}$  wavelength region, due to the  $\text{KrK}$  absorption edge. The otherwise much lower absorption of krypton is in part compensated by increasing the gas pressure to 50 cmHg.

Using a krypton-filled proportional counter for the detection of  $\text{MoK}\alpha$  X-rays brings to prominence a phenomenon which is of considerable importance for measurements requiring pulse height discrimination, viz. the "escape peak". This will be discussed in the section on pulse height discrimination, where the concept of "energy resolution" will also be introduced. Since this concept as well as the phenomenon of the escape peak are of like importance in the scintillation counter, the latter should first be described.

### The scintillation counter

The scintillation counter is composed of two basic elements: a flat fluorescent single crystal, and a photomultiplier tube (fig. 6). A NaI-crystal activated by Tl (1%) is commonly used for soft X-rays. An X-ray quantum absorbed by an atom of the crystal produces a photo-electron and usually

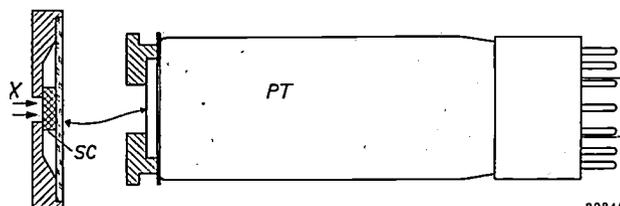


Fig. 6. Schematic representation of the scintillation counter, type 52245. The scintillating NaI:Tl-crystal (SC) is mounted in a light-tight holder, fitted with a glass cover transmitting the scintillation light to the end of the photomultiplier tube (PT) and an X-ray window (beryllium foil) at the other side.

one or more Auger electrons, which on their tracks energize a large number of fluorescent centres located at the sites of Tl ions in the crystal lattice. Each fluorescent center thus energized may emit a visible photon. The light so obtained is directed onto the cathode of the photomultiplier tube by reflecting surfaces near or on the crystal. For roughly 10 of these photons one electron is liberated from the cathode. The photomultiplier owing to its secondary electron emission gain of  $10^5$  or  $10^6$  delivers a current pulse which after amplification is suitable for operation of the counting circuits.

The scintillation counter as adapted to soft X-rays has characteristics similar to the proportional counter in two respects: it has an extremely short dead time (about  $0.2 \mu\text{sec}$ ), ensuring linear counting up to very high intensities; and its pulse amplitudes are proportional to the quantum energy, thus offering the possibility of pulse height discrimination. The outstanding feature of the scintillation counter as compared to both the Geiger and the proportional counter is its high quantum counting efficiency (nearly 100%) and its almost uniform spectral response in the wavelength region  $0.2$ - $2.5 \text{ \AA}$ . This is shown in fig. 7 for a crystal 1 mm thick. The quantum counting efficiency of nearly 100% is due to the practically complete absorption of the soft X-rays in the NaI-crystal slab<sup>10</sup>). Much thinner crystals of NaI could be used for X-ray analysis owing to its very high X-ray absorption coefficient. The area

<sup>10</sup>) The efficiency of scintillation counters in their "native" field, viz. nuclear physics, is rarely higher than say 1%, owing to the very incomplete absorption of the hard nuclear radiations in crystals of reasonable size.

of the crystal need be only slightly larger than the cross-section of the X-ray beam.

Other characteristics which make the NaI-crystal particularly suited for the purpose, are the following. Its spectral emission curve matches the sensitivity curve of photomultiplier tubes rather well; it has a high optical transparency; its fluorescence decay constant is very small, about  $10^{-7}$  sec, which accounts for the short dead time mentioned above; it is easily grown in single crystals of sufficient size and uniform transparency. The one important disadvantage of NaI:Tl is that it is extremely hygroscopic. The crystal therefore is hermetically sealed in an aluminum holder, which itself is completely opaque to external visible light. For good transparency to the incident X-rays, the front of the holder exposed to the beam is provided with a 0.13 mm thick beryllium window. The other side of the crystal mount has a flat glass plate and is "wrung" onto the photomultiplier tube face, with a fluid film of high refractive index (Dow Corning fluid 200) in order to minimize reflection losses. For the same reason the inner surfaces of holder and beryllium X-ray window are highly polished. Since beryllium is difficult to polish, the side of this window next to the crystal is sometimes covered with a very thin ( $\sim 1 \mu$ ) bright aluminum foil. (It should be noted that light losses do not affect the quantum counting efficiency, but do impair the energy resolution, see later.)

caused by the increasing absorption of the beryllium window. Windows of better transparency at wavelengths greater than  $3 \text{ \AA}$  are conceivable, but for the very soft X-rays the scintillation counter is difficult to use because the pulses become very small and it becomes difficult to distinguish them from the noise generated by the photomultiplier tube.

From this it will be clear that the photomultiplier tube should be designed for a very low noise ("dark current" caused e.g. by ohmic leakage, thermionic and field emission, gaseous ionization). Cooling the photomultiplier in order to reduce thermionic emission is not necessarily of benefit in reducing the noise, since in many photomultiplier tubes the noise pulses of largest amplitude, and hence the most troublesome ones, arise from gaseous ionization.

Other characteristics of the photomultiplier, such as the electron yield per incident photon and the total electron gain, have been mentioned above. Suitable types of photomultiplier tubes require anode voltages of 850 to 1400 V. The scintillation counter shown in fig. 1 is operated with an amplifier of gain 250 to 1000, depending on the voltage applied to the photomultiplier tube. The useful life of the scintillation counter is practically unlimited.

#### Pulse height discrimination

Pulse height discrimination as applied in X-ray diffractometry and X-ray spectrochemical analysis has the object of measuring radiation of a single wavelength, i.e. one specific diffraction line or one specific spectral emission line, while radiation of other wavelengths is eliminated<sup>11</sup>). The pulses delivered by a proportional or scintillation counter are fed via a linear amplifier (output pulse proportional to input pulse) with adjustable gain to a single-channel pulse height analyzer. This may be set to pass all pulses greater than a given amplitude (base) or only those pulses lying within a specified, rather narrow amplitude range (channel width or "window"). The pulses passed by the analyzer are fed to a recording counting-rate meter<sup>2</sup>).

A striking example illustrating the usefulness of pulse height discrimination is found in the diffractometry of radioactive samples<sup>12</sup>). Owing to the large size of the specimen used in diffractometry, the detector in the case of a radioactive sample is

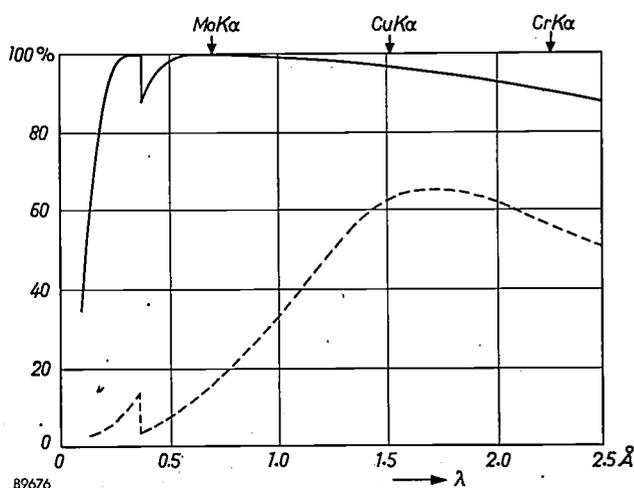


Fig. 7. Spectral response of the NaI:Tl scintillation counter type 52245, calculated from absorption data<sup>5</sup>). The dotted line is the response curve for the xenon-filled proportional counter from fig. 5.

The decrease of the spectral response curve of the scintillation counter in fig. 7 towards short wavelengths is due to the harder X-rays passing through the thin crystal without being absorbed. The decrease towards longer wavelength, however, is

<sup>11</sup>) The general methods of pulse-amplitude analysis and the apparatus used cannot be described in this article. See e.g.: A. B. van Rennes, Pulse-amplitude analysis in nuclear research, *Nucleonics* 10, July 1952, pp. 20-27, and August 1952, pp. 22-28. See also <sup>6</sup>).

<sup>12</sup>) D. J. Knowles, 10th Pittsburgh Conf. X-ray Electr. Diff. 1952, p. 32. T. R. Kohler and W. Parrish, X-ray diffractometry of radioactive samples, *Rev. sci. Instr.* 26, 374-379, 1955 (No. 4).

exposed to an intense background radiation. Shielding of the detector by lead (and scanning by moving the X-ray tube, with its heavy shield, instead of the detector) is a possible but cumbersome solution. If, however, a scintillation or proportional counter is used, the vast majority of pulses produced by the  $\beta$ - or  $\gamma$ -rays have amplitudes different from those produced by the diffracted X-rays (they may be larger as well as smaller); they can therefore be largely eliminated by a pulse height discriminator with the window set for the energy (wavelength) of the diffraction line whose intensity is to be measured.

#### Energy resolution of different detectors

When a given radiation is being measured while the upper and the lower amplitude limits of the analyzer window are continuously shifted in synchronism with the recording chart (keeping a constant window width), a graph of the pulse amplitude distribution of the pulses of the detector is obtained. Fig. 8 shows the pulse amplitude

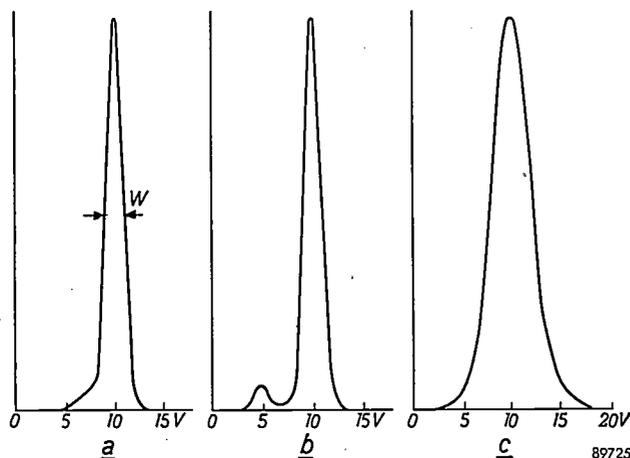


Fig. 8. Amplitude distribution curves of the pulses generated by  $\text{CuK}\alpha$  radiation in a) krypton-filled proportional counter, b) xenon-filled proportional counter, c) scintillation counter.

distribution curves, obtained with a window 0.5 V wide, for three types of detectors exposed to monochromatic radiation ( $\text{CuK}\alpha$ , which gives 10 V pulses). Each curve exhibits a peak of finite width, which means that even quanta of a single energy (i.e. 8 keV) produce pulses of different amplitudes, varying around an average value; this value corresponds to the true quantum energy. The variation, which limits the power of the detector to separate quanta of closely spaced wavelengths, is due to the statistics of the ionization or excitation and amplification processes involved. The separating power, or "energy resolution" (which of course cannot be enhanced by a narrower analyzer window) is measured by  $W/A$ ,  $W$  being the width at one half

peak height of the pulse amplitude distribution curve and  $A$  the average pulse amplitude (roughly the abscissa of the centre of the peak).

In fig. 8 it is seen that the energy resolution  $W/A$  for  $\text{CuK}\alpha$  radiation is much better for the proportional counter, filled with xenon or krypton, than for the scintillation counter. The better resolution of the proportional counter can be understood as follows. One  $\text{CuK}\alpha$  quantum will produce an average number  $m$  of 350 ion pairs in the xenon-filled proportional counter. Owing to fluctuations of  $m$  and of the gas amplification factor for each electron, the resulting pulse amplitude fluctuates with a relative standard deviation  $\sigma/A$ , which can be shown<sup>13)</sup> to amount to about  $1/\sqrt{m}$ . Using the relation  $W = 2.36 \sigma$ , the ratio  $W/A$  is calculated to be 13% for  $m = 350$ . The actual value measured in mass-produced counter tubes is about 20% (irregularities of the anode wire and slight asymmetry in its centering, which cause fluctuations of the gas amplification, probably account for the difference). In the scintillation counter, one  $\text{CuK}\alpha$  quantum will produce about 500 visible photons in the  $\text{NaI.Tl}$  crystal<sup>14)</sup>, but only about  $n = 25$  of these photons give rise to an "electron avalanche" in the photomultiplier tube. (The other photons are ineffectual because either they do not arrive at the photocathode, or they fail to liberate an electron, or the liberated electron fails to reach the first dynode, or because of fluctuations in the secondary emission process at the dynodes<sup>15)</sup>.) The relative standard deviation again is  $\sigma/A \approx 1/\sqrt{n} = 1/\sqrt{25}$ , so that  $W/A = 2.36 \sigma/A = 47\%$ . Thus the ratio  $W/A$  for the scintillation counter is in this example about 2.5 times larger and its energy resolution about 2.5 times smaller than that of the proportional counter. Significant improvements in the energy resolution of both these types of detectors do not appear likely at present.

From the relations  $W \propto 1/\sqrt{m}$  and  $W \propto 1/\sqrt{n}$  implied in the above it is seen that for both types of counter tubes  $W/A \propto 1/\sqrt{A}$ : the energy resolution improves with the square root of the quantum energy

<sup>13)</sup> U. Fano, Phys. Rev. 72, 26-29, 1947; O. R. Frisch (unpublished); S. C. Curran, A. L. Cockroft and J. Angus, Phil. Mag. 40, 929-937, 1949.

<sup>14)</sup> About 20% of the quantum energy is converted into visible light; the quantum wavelength is 1.54 Å, the average wavelength of the visible emission is 4100 Å, so that the number of visible photons per X-ray quantum is  $0.2 \times 4100/1.54 \approx 500$ . — It should perhaps be noted that the limited energy-conversion efficiency (20%) does not detract from the nearly 100% quantum counting efficiency of the scintillation counter for  $\text{CuK}\alpha$ , mentioned earlier: practically every absorbed quantum will be counted anyhow, independent of the loss of large part of its energy in the crystal.

<sup>15)</sup> G. A. Morton, RCA Rev. 10, 525-555, 1949.

of the detected radiation (i.e. — within obvious limits — inversely with the square root of its wavelength).

When the composition of the radiation to be measured and the behaviour of the detector are not known, the desirable setting of the pulse height analyzer window may be quickly determined by using only the base of the analyzer (i.e., with the upper limit of the window inoperative, see above) and making a recording with synchronized base shift and chart movement. This will yield what is called the "integral curve" (number of pulses per second having an amplitude exceeding the value on the axis of the abscissa). The integral curves obtained for CuK $\alpha$  radiation with the same three detectors as in fig. 8 are reproduced in fig. 9.

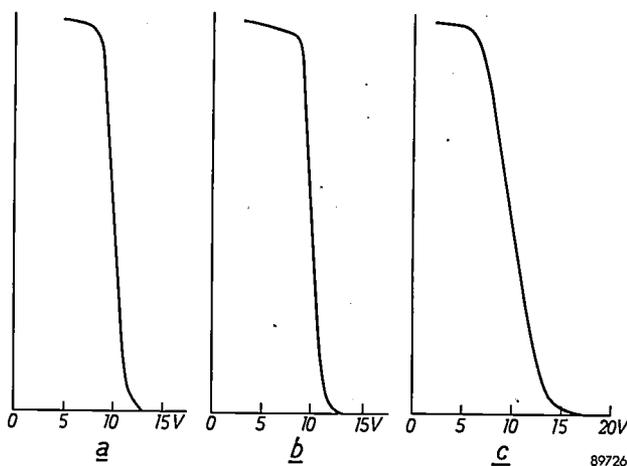


Fig. 9. Integral curves of the same pulses as in fig. 8. The ordinate of each curve indicates the number of pulses per second having an amplitude exceeding the voltage value plotted as abscissa.

#### The escape peak

The primary process in the absorption of an X-ray quantum by all the counter tubes considered is the ejection of an electron from an inner shell of a gas or crystal atom. This photo-electron will originate say from the K shell, if the wavelength of the absorbed quantum is shorter than that of the K absorption edge of the atom. The following transition of an L shell electron to the K shell vacancy is associated with either the ejection of one or more other electrons (Auger process) or the emission of a new X-ray quantum (fluorescence). It has been stated above that the Auger electrons similarly to the photo-electron will ultimately be absorbed in the gas, or crystal, and will contribute to the formation of ion pairs, or to the excitation of fluorescent centres. The alternative fluorescent X-ray quantum, however, has a sizable probability of *escaping* from the gas or crystal, especially since its wavelength

is always somewhat longer than that of the absorption edge and absorption for these quanta is therefore relatively weak (cf. figs. 5 and 7). For the Geiger counter it does not matter greatly whether a substantial proportion of the original energy  $E_i$  of the quantum manages to escape as fluorescent energy,  $E_f$ . For the proportional and scintillation counter, however, the escaping of fluorescent X-radiation has the undesirable consequence that some of the original quanta result in pulses with a "wrong" amplitude, viz. proportional to  $E_i - E_f$  instead of  $E_i$ . Thus two peaks appear in the pulse distribution curve for monochromatic radiation, the main peak (at  $V_i \propto E_i$ ) and the "escape peak" (at  $V_e \propto E_i - E_f$ ).

A small escape peak is visible in fig. 8b for CuK $\alpha$  radiation incident on the xenon-filled proportional counter. A very large escape peak — larger than the main peak — is found when MoK $\alpha$  radiation ( $E_i = 17.4$  keV) is measured with a proportional counter filled with krypton (the escaping fluorescent KrK $\alpha$  radiation has  $E_f = 12.6$  keV). This is illustrated by fig. 10.

The existence of the escape peak will influence practical measurements with pulse height discrimination in two ways. Since the analyzer window usually has to be made so narrow that it will reject pulses of the escape peak, the quantum counting efficiency will be correspondingly reduced. The loss would amount to about 55% in the case of fig. 10. Secondly, for quanta  $E_i + E_f$ , if present, the escape peak would inevitably be passed by the analyzer window set for the main peak of quanta  $E_i$ ; the discrimination against unwanted radiation (e.g. a continuous background) is therefore incomplete.

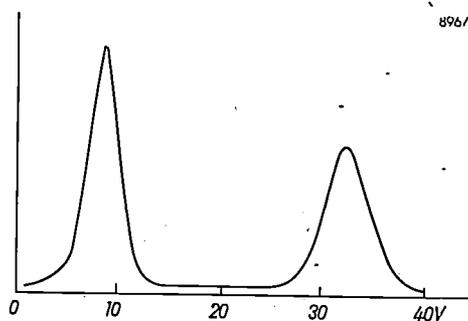


Fig. 10. Amplitude distribution of the pulses produced in the krypton-filled proportional counter by MoK $\alpha$  radiation (quantum energy  $E_i = 17.4$  keV). The pulses forming the "main peak" have amplitudes (average 32 V) corresponding to the ionization produced by both photo-electron and Auger electron(s). The amplitudes of the other group of pulses (average 9 V), forming the "escape peak", correspond to the residual ionization when fluorescent X-rays are lost. The amplitude difference (23 V) between pulses of main peak and escape peak is determined by the energy  $E_f$  of the escaped fluorescent X-ray quantum (KrK $\alpha$ , with  $E_f = 12.6$  keV) and therefore remains constant when the wavelength of the detected radiation is varied.

In some cases the escape peak can be put to good use, since the energy resolution of the counter is better in this part of the pulse amplitude distribution curve than in the main peak. This can be seen as follows. Consider the problem of separating, in spectrochemical analysis, two closely spaced emission lines, of quantum energies  $E_{i1}$  and  $E_{i2}$ , and assume that the detector delivers an escape peak for each line, the energy of the escaping fluorescent quantum being  $E_f$ . The mean pulse amplitude  $A_e$  of the escape peak in each case is smaller than that of the main peak  $A_m$  by the same amount  $\Delta A$ , proportional to  $E_f$ . Thus the escape peaks of both radiations have the same separation as the main peaks,  $A_{e1} - A_{e2} = A_{m1} - A_{m2}$ . The half height width  $W$  of any peak in the pulse amplitude distribution, however, is proportional to the square root of the mean amplitude,  $\sqrt{A}$  (see above). Each escape peak, therefore, is narrower than the main peak by a factor  $\sqrt{A_e/A_m}$ . Evidently the narrower escape peaks with equal separation can be more easily resolved than the main peaks.

quantum counting efficiency of the proportional counter is low at these short wavelengths.

Incidentally, it is seen in this recording that one radiation may produce several escape peaks. In fact, besides  $K\alpha$  fluorescent quanta,  $K\beta$  quanta and in other cases  $L$  quanta may arise, with their corresponding fine structure. Usually these niceties can be disregarded.

### Comparison of the three types of detectors in practical applications

When using the "Norelco" X-ray instrumentation, it is necessary to choose for each problem the most suitable detector among the three available types. The considerations determining this choice will now be discussed. In doing so, it seems logical to start with X-ray spectrochemical analysis: shortcomings

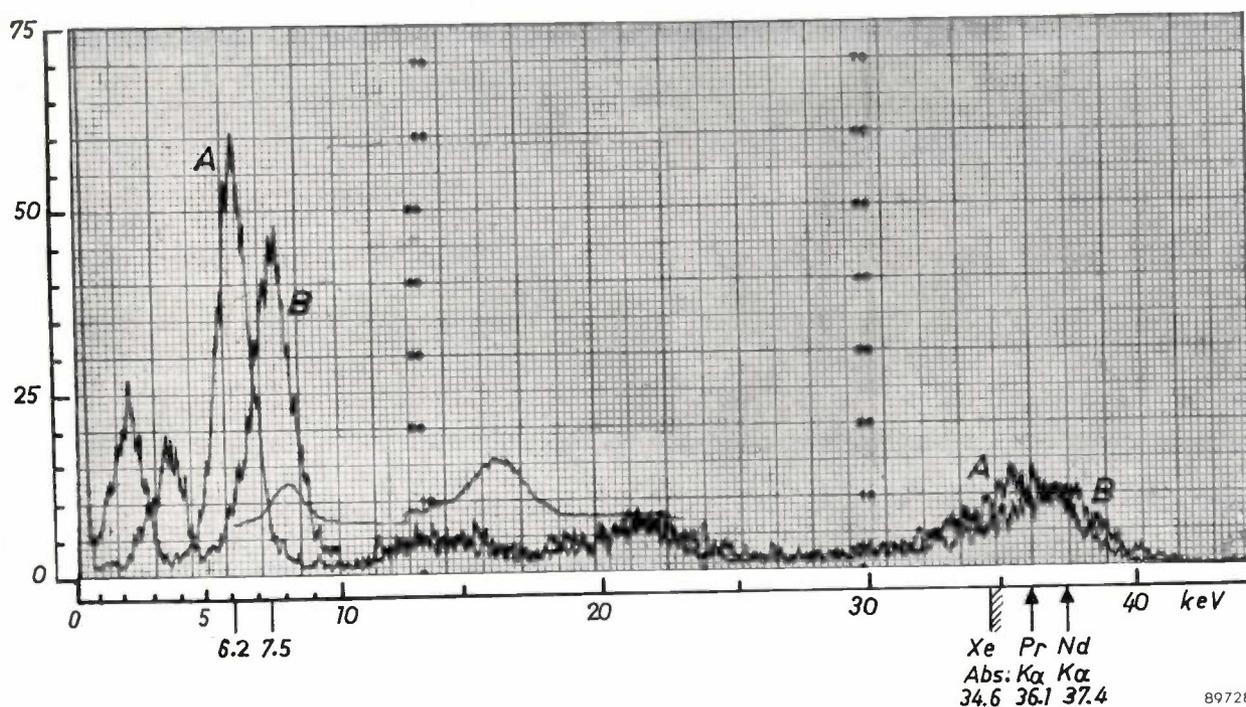


Fig. 11. Illustration of the fact that the resolving power of a counter is better in the escape peak region than in the main peak region. The pulse amplitude distribution curves *A* and *B* were obtained from a xenon-filled proportional counter with monochromatic radiations of energies 36.1 keV (praseodymium  $K\alpha$ ) and 37.4 keV (neodymium  $K\alpha$ ) respectively.

Use can be made of this fact by setting the window of the pulse height analyzer for one of the two escape peaks rather than for one of the two main peaks.

A striking example is seen in *fig. 11*, which shows the pulse amplitude distribution curves of a xenon-filled proportional counter for the  $K\alpha$  radiations of praseodymium (atomic number 59,  $E_i = 36.1$  keV) and neodymium (atomic number 60,  $E_i = 37.4$  keV). Both radiations have wavelengths shorter than the K-absorption edge of xenon and produce a very strong escape peak (85% of the total number of output pulses are grouped in this peak!). Since  $E_f = 28.9$  keV for Xe  $K\alpha$ , it is calculated that the escape peak width is about 0.5 times the main peak width. It should be noted, however, that the

of the Geiger counter for this particular application gave the fillip to the introduction of proportional and scintillation counters into the X-ray field, so that here the case is rather clear-cut.

### Spectrochemical analysis<sup>3)</sup>

In spectrochemical analysis, radiations of *different* wavelengths, whose intensities may vary greatly, must be compared quantitatively or semi-quantitatively. The non-linearity of the Geiger counter due to its long dead time is a serious dis-

advantage in this case, since it will generally necessitate a correction of the measured peak intensities. Such a correction is neither easy nor accurate, particularly because the operational dead time of the counter when the X-ray tube is not supplied with a constant potential but e.g. with a full-wave rectified voltage, will depend on the wavelength<sup>16</sup>). For intensities exceeding a few thousand counts/sec (intensities exceeding  $10^5$  counts/sec may occur), corrections are no longer possible and it will be necessary to reduce the intensities to the Geiger counter range by controlled absorption. This becomes very difficult for a series of different wavelengths.

Both the proportional and the scintillation counter have much better linearity than the Geiger counter. Their dead time ( $\sim 0.2$   $\mu$ sec) would cause non-linearity only at  $10^7$  counts/sec or even higher. The practical limits of linearity in this case are set by the associated electronic circuits. The standard "Norelco" scaling circuit gives rise to a count loss of 10% at 14000 counts/sec. New circuits now available are linear to 1% at 10000 counts/sec. Similarly the pulse height analyzer hitherto used in our investigations precluded the measurement of counting rates higher than about 5000 counts/sec, owing to a shift of the base level depending on the counting rate. Better analyzers are now available which retain the simplicity of the present instrument but permit intensity measurements up to about 15000 counts/sec or even higher, depending upon the counting rate of the pulses being rejected by the analyzer.

The strongly wavelength-dependent quantum counting efficiency, which the proportional counter shares with the Geiger counter (fig. 7), requires corrections to the measured intensities when characteristic lines of elements several atomic numbers apart are compared and no reference standards are used, as is frequently done in qualitative and semi-quantitative analysis. Although there are several other factors which influence the relative line intensities, it is an important advantage of the scintillation counter for spectrochemical analysis that owing to its nearly uniform spectral response, corrections for variations in the efficiency are not necessary. Even more important is its very high value of the quantum counting efficiency, particularly for wavelengths  $< 1$  Å. This advantage is directly reflected in the time necessary for each measurement, since a fixed

number of counts must be accumulated for a given accuracy (cf. <sup>2</sup>); the counting rate obtained with the scintillation counter exceeds that of the xenon-filled proportional counter by a factor of almost 2 for CrK $\alpha$  and of about 12 for AgK $\alpha$ .

For long wavelengths the scintillation counter loses its advantage owing to the rather large photomultiplier noise mentioned previously. Measurements with all types of counters are hampered here by the difficulty of making sufficiently transparent X-ray windows. Using non-vacuum-tight windows in special flow counters has extended the useful range of the proportional counter with its inherent low noise to about 12 Å (in a helium path goniometer, cf. <sup>3</sup>): K $\alpha$  lines of silicon (7.13 Å), aluminum (8.34 Å), magnesium (9.89 Å) and sodium (11.91 Å) could thus be measured with a good peak-to-background ratio<sup>17</sup>), whereas the longest wavelength for which we were able to make useful measurements with the scintillation counter, using a specially selected crystal and photomultiplier tube, was K $\alpha$  of sulphur (5.37 Å). TiK $\alpha$  (2.8 Å) is about the longest wavelength that can be measured in routine fashion with the standard "Norelco" scintillation counters.

The possibility of pulse height discrimination with the scintillation counter and proportional counter is a great asset in spectrochemical analysis when peak intensities become smaller. Thus the analysis of minor and trace amounts of elements is achieved with better accuracy and lower limits of detectability. The usefulness of pulse height discrimination for separating overlapping spectral lines of different elements has been discussed in a previous article<sup>3</sup>), as well as the possibility of a non-dispersive spectrochemical analysis based on those principles. Clearly the greater energy resolution of the proportional counter may be advantageous in these cases, although one must balance this advantage against the higher quantum counting efficiency of the scintillation counter for short wavelengths.

#### *Powder diffractometry*<sup>1)</sup>

The situation in diffractometry is quite different from that in spectrochemical analysis, for two reasons: the intensities to be measured are usually much lower, and in principle use is made of monochromatic radiation, e.g. CuK $\alpha$  (1.54 Å) or MoK $\alpha$  (0.71 Å) or CrK $\alpha$  (2.29 Å). The Bragg reflections of this radiation obtained from the specimen (sharp

<sup>16</sup>) For a given peak voltage on the X-ray tube, the fraction of the a.c. cycle during which the elements in the specimen can fluoresce and into which, therefore, the quanta to be counted are bunched, decreases as the atomic number increases. Thus the effective dead time becomes longer for regions of shorter wavelengths.

<sup>17</sup>) C. F. Hendee, S. Fine and W. Brown, *Rev. sci. Instr.* 27, 531-535, 1956 (No. 7). Prof. C. H. Shaw of Ohio State University, according to a personal communication, has reached about the same limit by using proportional counters with a beryllium foil window.

lines) are superimposed on a background consisting chiefly of two contributions: radiation of a continuous spectrum, which is unavoidably emitted by the X-ray tube and scattered by the specimen, and characteristic fluorescent X-radiation excited in the specimen<sup>18</sup>). The scattered continuous background has its maximum intensity at wavelengths usually much shorter than the useful radiation. A fluorescent background — which will appear only if the specimen contains elements whose excitation voltage is sufficiently low — has wavelengths shorter or longer than the useful radiation. The total intensity and the spectral distribution of the background thus may vary widely, depending on the specimen, the X-ray tube voltage and other experimental conditions.

Counter tubes for powder diffractometry should have a high quantum counting efficiency, because of the low intensity of most diffraction lines, and should yield a high peak-to-background ratio, for the measurement of weak lines. Linearity is usually of less importance, and so is the uniformity of the spectral response. The non-uniform spectral response of the Geiger counter (and proportional counter) in fact may be regarded as an advantage for some applications of powder diffractometry, since it produces a crude sort of monochromating effect, which eliminates a large part of the short-wavelength

continuous scattered background. A qualitative conclusion is that for many applications of diffractometry, especially when simplicity of equipment matters, the Geiger counter offers a very satisfactory solution. Nevertheless, the scintillation counter, using more elaborate equipment, is capable of even better peak-to-background ratios: the lack of the above-mentioned monochromating effect is overcome by pulse height discrimination. The high quantum counting efficiency of the scintillation counter is another advantage. This efficiency, on the other hand, is diminished when pulse height discrimination is applied: the pulse height analyzer window, set for the peak in the pulse amplitude distribution curve for the monochromatic radiation used (fig. 8), will transmit only a certain percentage of the relevant pulses owing to the finite peak width. The quantum counting efficiency multiplied by the "transmission factor", will be called the "detection efficiency". Making the window wider will increase the transmission factor and therefore the detection efficiency, but reduce the monochromating effect.

From the above it will be clear that for choosing the best detector for any particular application of powder diffractometry, a quantitative comparison of the detection efficiency and the peak-to-background ratio obtained with each of the three types of detectors is necessary. Giving equal importance to these two quantities, their product may be calculated and used as a "figure of merit" of the detector. A large number of measurements have been made in order to establish these figures of merit for a variety of cases<sup>6</sup>).

As a first instance, a typical inorganic specimen was selected and measurements were made using the three radiations previously mentioned above.

<sup>18</sup>) It is assumed that non-specimen scatter, due to misalignment of the goniometer, is practically eliminated. The contribution of air scatter can usually be neglected. In addition to the continuous radiation, a portion of the useful monochromatic radiation is scattered (not "reflected") by the specimen. This contribution to the background of course cannot be distinguished from the diffraction line proper by pulse height discrimination. — A possible non-X-ray background, as in the diffractometry of radioactive samples<sup>12</sup>), need not be considered here.

Table I. Comparison of detectors for powder diffractometry. Measurements performed on the (111) reflection of a pure silicon powder specimen (no fluorescent background). For each of the three X-ray tubes (target Mo, Cu and Cr respectively) the voltage was empirically selected so as to give the highest ratio of useful characteristic to undesired "white" radiation. The pulse height analyzer window was set for a transmission factor of about 0.90. The Geiger counter measurements were corrected for dead time losses ( $\tau = 270$   $\mu$ sec). The figures of merit are relative values, with the first in each group arbitrarily taken as 100.

Detector	MoK $\alpha$ , 55 kV <sub>p</sub> , 7 mA; filter Zr 0.002"; peak P at 2 $\theta$ = 13.0°; background B at 10.0°; beam apert. 2 $\alpha$ = 0.5°			CuK $\alpha$ , 40 kV <sub>p</sub> , 9.6 mA; filter Ni 0.0007"; peak P at 2 $\theta$ = 28.4°; background B at 20.0°; beam apert. 2 $\alpha$ = 1.0°			CrK $\alpha$ , 30 kV <sub>p</sub> , 12.8 mA; filter V <sub>2</sub> O <sub>5</sub> ; peak P at 2 $\theta$ = 42.9°; background B at 35.0°; beam apert. 2 $\alpha$ = 2.0°		
	Det. eff. %	P/B	Figure of merit	Det. eff. %	P/B	Figure of merit	Det. eff. %	P/B	Figure of merit
Scintillation counter with pulse height discrimination	90	44	100	89	134	100	90	93	100
Scint. counter without pulse ht. discr.	100	10	25	100	12	10	100	5	6
Prop.c. (Xe) with pulse ht. discr.	12	44	13	56	146	69	58	92	64
Prop.c. (Xe) without pulse ht. discr.	13	16	5	63	57	30	60	16	11
Prop.c. (Kr) with pulse ht. discr.	12	51	15	33	64	18	—	—	—
Prop.c. (Kr) without pulse ht. discr.	29	29	21	36	26	8	—	—	—
Geiger counter	14	27	10	51	46	20	59	18	13

The specimen gives no measurable fluorescence with any of these radiations. The results are listed in Table I. For MoK $\alpha$  the superiority of the scintillation counter over the other detectors is very marked; this evidently is due to its much higher quantum counting efficiency at this wavelength. The figure of merit of the krypton-filled proportional counter is seen to be lowered when pulse height discrimination is applied. This fact is caused by the strong escape peak, whose suppression reduces the detection efficiency more than is gained in the peak-to-background ratio. For CuK $\alpha$  and CrK $\alpha$  the scintillation counter with pulse height discrimination also

low for CuK $\alpha$  and CrK $\alpha$ . The effectiveness of pulse height discrimination in the latter case is very clearly illustrated by fig. 12.

It is interesting to note that the peak-to-background ratios obtained with the scintillation and Xe proportional counter are several times higher for CuK $\alpha$  and CrK $\alpha$  than for MoK $\alpha$ . The reason for this is that the characteristic MoK $\alpha$  line is nearly at the peak of the continuous spectrum, whereas the CuK $\alpha$  and CrK $\alpha$  lines are well removed from the peak<sup>19</sup>.

A second series of measurements, which will not be reported here in detail, was concerned with the scattered backgrounds in the diffractometer patterns

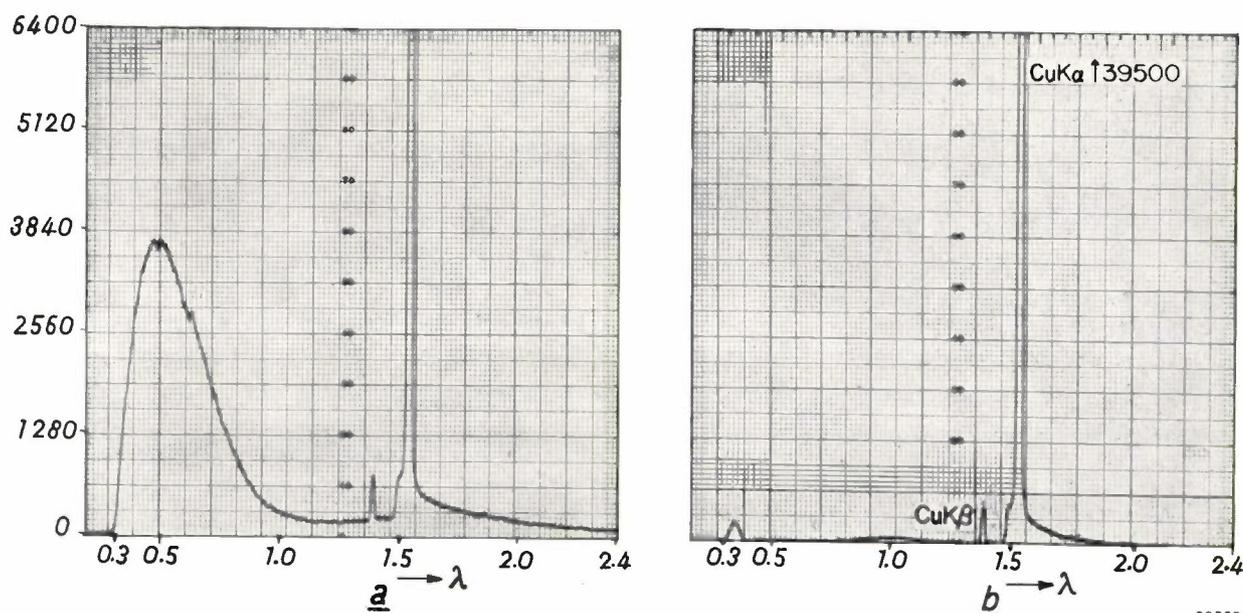


Fig. 12. Spectrum from Cu-target X-ray tube (40 kV<sub>p</sub>, full-wave rectification) obtained with scintillation counter and silicon ((111) plane) analyzer crystal, *a*) without, *b*) with pulse height discrimination, the window being set for CuK $\alpha$  with a transmission factor of 0.90. A Ni-filter 0.0007" thick, which virtually eliminates the CuK $\beta$  line and also reduces the continuum, was employed in both cases. The nearly complete suppression of the continuous spectrum in (*b*) is responsible for the enormous improvement in the peak-to-background ratio in diffractometer measurements of polycrystalline specimens with CuK $\alpha$ , as shown in Table I (increase from 12 to 134). Incidentally, the small residual hump at about 0.35 Å is due to the escape peak phenomenon: radiation of wavelength  $\sim$  0.35 Å owing to this phenomenon will produce a certain proportion of pulses falling within the analyzer window set for 1.54 Å.

rates as the best detector but not by so wide a margin as in the case of MoK $\alpha$ . When pulse height discrimination is not applied, the advantage of the Geiger counter (and of the Xe proportional counter) in not being sensitive to the short-wavelength scattered radiation is clearly seen; the Kr proportional counter has a very low figure of merit in this case owing to the rise of its quantum counting efficiency at the KrK absorption edge (fig. 5), and for similar reasons the figure of merit of the scintillation counter *without* pulse height discrimination is

of different (non-fluorescent) specimens, measured at several diffraction angles. The total background without pulse height discrimination varied by a factor of 20 among the specimens tested. The reduction of total background by the use of pulse height discrimination varied from a factor of about 2 (specimen hexamethylene tetramine) to 53 (specimen silver), for the scintillation counter<sup>6</sup>). Dis-

<sup>19</sup> W. Parrish and T. R. Kohler, A comparison of X-ray wavelengths for powder diffractometry, *J. appl. Phys.* **27**, 1215-1218, 1956 (No. 10).

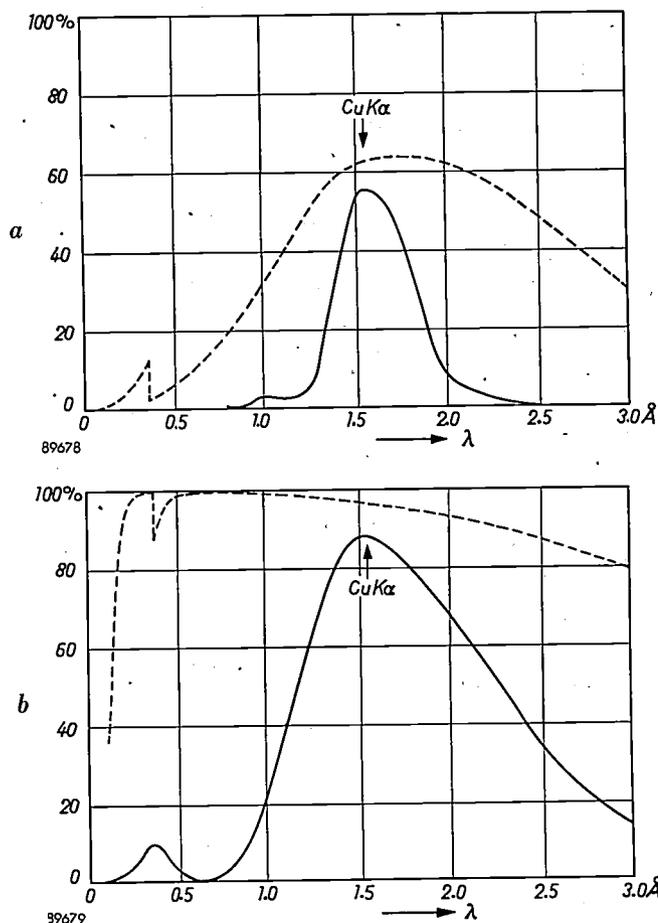


Fig. 13. Detection efficiency for monochromatic radiations as a function of  $\lambda$ , with symmetrical pulse height analyzer window (transmission factor 0.90) set for  $\text{CuK}\alpha$ . a) Proportional counter filled with xenon. b) Scintillation counter. The data are calculated from the quantum counting efficiencies (dotted curves, as in fig. 7) and the pulse amplitude distribution curves.

crimination with the Xe proportional counter gives a smaller reduction factor, as may be expected from its lower quantum counting efficiency in the short wavelength region.

Finally a number of measurements were made concerning the background due to specimen fluorescence. The diffraction pattern of iron powder obtained with  $\text{CuK}\alpha$  radiation was selected as a rather difficult case. The background from the

lower level, resulting in an *asymmetrical* window will decrease the detection efficiency for  $\text{FeK}\alpha$  by a larger factor than that for  $\text{CuK}\alpha$ . A ratio of 25 is obtained with the asymmetrical window indicated iron specimen contains strong  $\text{FeK}\alpha$  and  $\text{FeK}\beta$  fluorescence, in addition to a certain amount of scattered short wavelength continuum. The  $\text{FeK}\alpha$  wavelength (atomic number 26,  $\lambda = 1.93 \text{ \AA}$ , energy 6.4 keV) is rather near that of  $\text{CuK}\alpha$  (atomic number 29,  $\lambda = 1.54 \text{ \AA}$ , energy 8.0 keV). The efficiency of pulse height discrimination in suppressing the undesired  $\text{FeK}$  radiation when the window is set for  $\text{CuK}\alpha$  will in this case largely depend on the energy resolution of the counter. Quantitative evaluation of the effect of discrimination is simpler for such a case than for the suppression of a continuous background, since the quantum counting efficiencies for only two wavelengths need be considered. The ratio of the detection efficiencies for these two wavelengths (i.e.  $\text{CuK}\alpha/\text{FeK}\alpha$ ) takes the place of the previously introduced peak-to-background ratio, and the figure of merit is redefined for this case as the product of this ratio times the detection efficiency for  $\text{CuK}\alpha$ .

The detection efficiency of a counter as a function of wavelength for a given setting of the analyzer window is easily calculated from its spectral response curve (fig. 7) and the pulse amplitude distribution curve for monochromatic radiation (for example fig. 8). The result is shown in fig. 13a for the xenon-filled proportional counter, in fig. 13b for the scintillation counter. In both cases a symmetrical window setting with a transmission factor of 0.90 was used.

The result looks rather disappointing: a  $\text{CuK}\alpha/\text{FeK}\alpha$  ratio of not more than 1.2 is obtained with the scintillation counter and of 1.6 with the proportional counter. It is not essential, however, to use the above-mentioned *symmetrical* setting of the pulse height analyzer window. Consider the integral curves of the proportional counter for  $\text{CuK}\alpha$  and for  $\text{FeK}\alpha$  radiation, fig. 14. The symmetrical window setting is indicated by vertical lines; a higher setting of the

Table II. Reduction of fluorescent background with pulse height discrimination (iron specimen, irradiated by radiation of copper target X-ray tube operated at 40 kV(peak); other experimental conditions see fig. 14).

	Symmetrical analyzer window		Asymmetrical analyzer window	
	Ratio of detection efficiencies $\text{CuK}\alpha/\text{FeK}$	Figure of merit	Ratio of detection efficiencies $\text{CuK}\alpha/\text{FeK}$	Figure of merit
Xe proportional counter	1.6	13.5	25	100
Scintillation counter	1.2	16.3	4.1	27.5

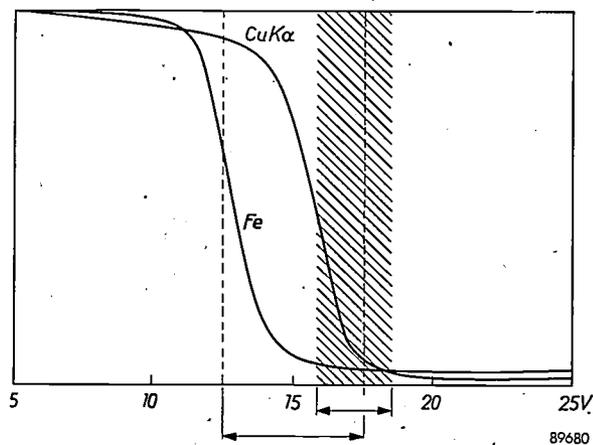


Fig. 14

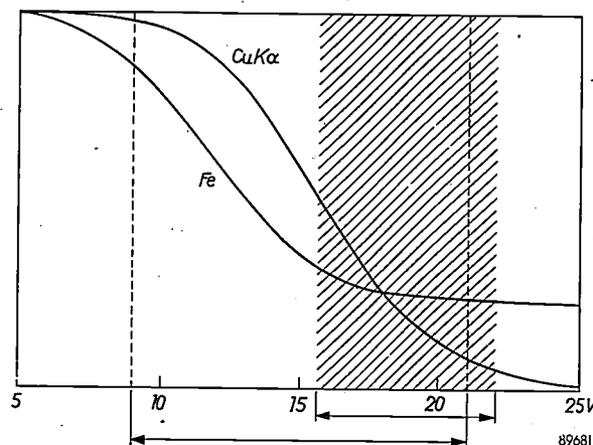


Fig. 15

Fig. 14. Integral curves of the pulses obtained from the xenon-filled proportional counter with  $\text{CuK}\alpha$  radiation (diffracted (111) peak of a silicon powder specimen) and with background of a pure iron specimen at  $2\theta = 23^\circ$ ;  $\text{FeK}$  fluorescence is predominant in this case. The vertical lines indicate the conventional symmetrical setting of the pulse height analyzer window (transmission factor 0.90). Experimental conditions: copper target X-ray tube, 0.0008" nickel filter for suppression of  $\text{CuK}\beta$ , beam aperture  $2\alpha = 1^\circ$ , receiving slit width 0.024". The tube was operated at 20 kV(peak) for silicon and at 40 kV(peak) for iron. The curves were normalized to have the same counting rate at 5 V pulse amplitude (this is permissible since only the relative attenuations caused by the analyzer are considered)

Fig. 15. Same as fig. 14, but for the scintillation counter. The smaller slope of the curves as compared with fig. 14 is due to the poorer energy resolution of the scintillation counter; the higher intensity level at pulse amplitudes above about 18 V (scattered short wavelength continuum) is due to the higher quantum counting efficiency of this counter at short wavelengths.

lower level, resulting in an *asymmetrical* window will decrease the detection efficiency for  $\text{FeK}\alpha$  by a larger factor than that for  $\text{CuK}\alpha$ . A ratio of 25 is obtained with the asymmetrical window indicated by shading in fig. 14. A similar but less spectacular improvement is obtained when an asymmetrical window is used for the scintillation counter, as shown in fig. 15. The results are summarized in Table II.

The smaller improvement with the scintillation counter is a consequence of its poorer energy resolution. On the other hand, even the better resolution of the proportional counter may prove inadequate when smaller wavelength differences

occur, as with  $\text{Co K}$  fluorescence (atomic number 27) excited by  $\text{CuK}\alpha$  radiation. In that case the use of a crystal monochromator placed after the specimen would be necessary in order to remove all the fluorescent background; the pulse height analyzer may then be omitted and because of its simplicity a Geiger counter will preferably be used for the detection of the monochromatic radiation, especially since the intensities in this case will be rather low. This example is perhaps a fitting conclusion for this article, since it reintroduces the eldest member of the family of X-ray counters and clearly illustrates that each member has specific applications in which it is superior to the others.

**Summary.** A Geiger counter, a xenon- or krypton-filled proportional counter and a scintillation counter are available as interchangeable detectors for the "Norelco" X-ray goniometer. The design and counting mechanism of each of these detectors is briefly described in this article. The main characteristics of the three detection methods are: the Geiger counter has large pulse amplitude and concurrent simplicity of equipment; the proportional and scintillation counters have linear response up to very high counting rates, and proportionality between pulse amplitude and quantum energy, offering the possibility of pulse height discrimination; moreover the scintillation counter has the highest quantum counting efficiency and a very uniform spectral response, whereas the proportional counters have a better energy resolution. After a discussion of some aspects of pulse height discrimination, especially energy resolution and the escape peak phenomenon, an attempt is made to evaluate the relative merits of the different types of detectors for several applications. For spectrochemical analysis,

the scintillation counter will in most cases be preferred for wavelengths up to about 3 Å. For longer wavelengths a very thin window proportional counter (either sealed off or flow type) is desirable. For powder diffractometry, the equal importance of high sensitivity and good peak-to-background ratio leads to the proposal of a figure of merit, which is calculated for different counters and for a typical non-fluorescing specimen analyzed with three radiations ( $\text{CuK}\alpha$ ,  $\text{MoK}\alpha$  and  $\text{CrK}\alpha$ ). The scintillation counter used in conjunction with pulse height discrimination has by far the highest figure of merit for  $\text{MoK}\alpha$  radiation; its superiority above the xenon-filled proportional counter is less pronounced for  $\text{CuK}\alpha$  and  $\text{CrK}\alpha$  radiation. Finally, a fluorescent specimen (iron powder analyzed with  $\text{CuK}\alpha$ ) is considered. Pulse height discrimination in this case is especially effective when an asymmetrical setting of the analyzer window is used, and the xenon-filled proportional counter is markedly superior to the scintillation counter owing to its better energy resolution.

## A MUSICAL INSTRUMENT FOR DEAF-MUTE CHILDREN

789.983:616.28-008.15

Deaf-mute children usually have only very little left of their hearing faculty. Nevertheless, they can perceive sounds of an intensity above a certain threshold value, either through their ears or as vibrations felt by any part of the body, particularly the thorax. If the children grow up as "deaf-mutes", i.e. have not learned to speak and are consequently retarded in their mental development, the reason is largely that they do not realize without instruction that the vibrations they perceive are capable of containing information.

By exercising the sense of vibration in the body it is frequently possible to awaken in deaf-mute children an awareness of the world of sound. It is worth while exploiting these possibilities to the full, for anything that can be done to enable a deaf-mute child to acquire impressions of sound, however imperfect, will contribute to its spiritual development and help it to overcome the almost insuperable difficulties of becoming a "normal" child.

One of the numerous methods at present being employed for training deaf-mute children to exercise their sense of vibration is to play music to them and, more important still, to let them make music themselves. A wind instrument is the most suitable for this purpose since it can also help the children to learn breath-control. Persons able to hear and speak normally are generally unaware of how fundamental this factor is in speech. One of the most important tasks in giving speech instruction to deaf-mute children is to teach them how to control their breathing<sup>1)</sup>.

About 15 years ago, some pupils of the Institute for the Deaf at St.-Michielsgestel in Holland were taught to play the clarinet, the vibrations of which can be felt both between the lips and in the chest. It was found that the children began spontaneously to hum the melody they had learned to play on the clarinet<sup>2)</sup>. However, the intensity with which they

could hear themselves playing this instrument was too weak to lead to sound perception, and moreover very young children, and older ones too, find the clarinet too difficult to play. For this reason the Institute requested our laboratory to develop a wind instrument that would be easy to play and that would enable the children to feel (and, where possible, to hear) the music with their ears by means of headphones and also to feel it over their bodies by means of loudspeakers.

The instrument developed in response to this request is shown in the adjoining photograph<sup>3)</sup>. It is a wind instrument but the notes are selected on a keyboard and reproduced electronically. Rather as in a mouth organ, the pitch of the notes is determined by a row of steel reeds mounted on aluminium housings. The pupil blows down a rubber pipe with a glass mouthpiece, and sets the reeds vibrating by depressing keys which open corresponding valves. In the enclosed space containing the reeds a piezo-electric microphone picks up the vibrations, which are then reproduced at the required level via an amplifier and a loudspeaker. The acoustic screening of the microphone was found to be sufficient to prevent acoustic feedback.

Thus, as with the clarinet and other woodwinds, the notes are formed by depressing keys, but now only one finger is needed and a simple keyboard is provided. The intensity of the notes, their start and duration, and thus also the rhythm, are controlled by the breath. In this connection the shape of the mouthpiece is thought to be of particular importance. For this reason (apart from hygienic considerations) the mouthpieces are interchangeable; a broad flange determines the correct position in the mouth.

The problem of condensation and saliva encountered in the playing of wind instruments was solved in this case by Hoeymans<sup>3)</sup> with a simple bellows arrangement, which works as follows. An enclosed space is divided into two chambers by a light, moveable flap, sealed at the edges by thin leather bellows. When air is blown into the one chamber, a nearly equal amount of air is forced out of the other by the displaced flap and this air sets the reed vibrating via the valve opened by the depression of a key. When the player stops blowing, a valve in the second (dry) chamber opens, allowing air from outside to be sucked in by the flap as it re-

<sup>1)</sup> Experiments by S. Woldring of the Philips Research Laboratory, Eindhoven, shortly to be published in the "Ned. Tijdschrift voor Doofstommenonderwijs" (Netherlands Journal on the Teaching of Deaf-Mutes) have made this abundantly clear. To read seven short sentences, two normal children had to take 4 and 6 breaths respectively, whereas three deaf children drew respectively 26, 20 and 40 breaths. The breath curves given in the above paper show the unchecked, irregular nature of the deaf children's breathing.

<sup>2)</sup> A. van Uden, Annual Reports of the Institute for the Deaf, St.-Michielsgestel, e.g. 1951, p. 60. For a more general introduction see also: A. van Uden, Rhythmic training in sound perception of severely deaf children, Conf. Nat. Coll. Teachers of the Deaf of Great Britain, Blackpool, 13 April 1955.

<sup>3)</sup> The instrument has been further developed and a number of them built by an Eindhoven music teacher, Mr. A. Hoeymans.

turns to its equilibrium position under the action of a weak spring. This valve closes automatically as soon as air is blown in again, whereupon the displaced air again passes to the reeds. Although it is not possible with this arrangement to produce a long

simply by feeling and hearing the melodies. After respectively  $3\frac{1}{2}$  and 4 hours of practice, the first group was able to recognize the tunes to more than 90%, the other only to 67%; it made hardly any difference in either case whether the children listened



sustained note, owing to the limited stroke of the bellows, Hoeymans has nevertheless evolved a design such that this does not prove troublesome in normal use.

The instrument, which has been in use for several years at the St.-Michielsgestel Institute, serves for individual tuition as well as for class instruction (see photograph overleaf). In the latter case the children can also hear (or feel) and join in the music played by the others.

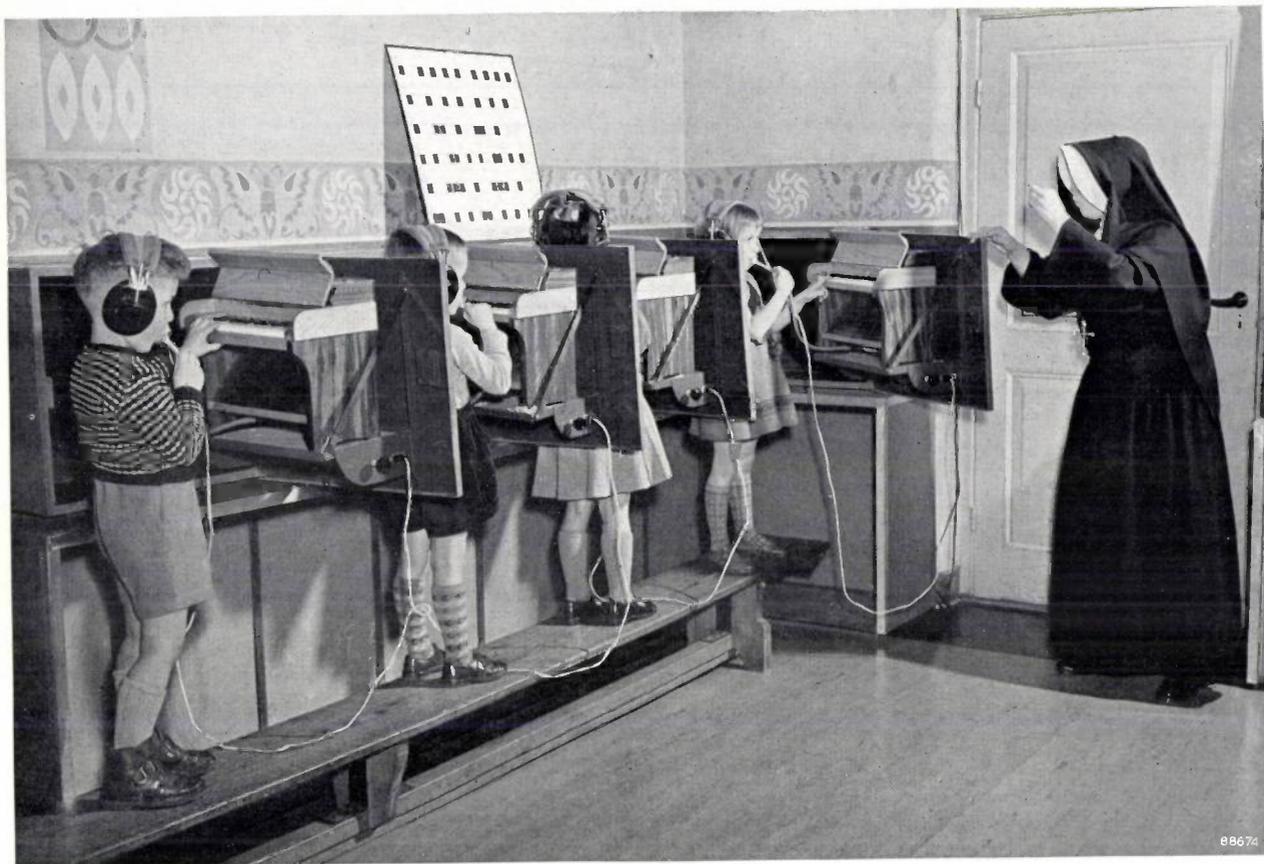
Some idea of the practical results obtained with this instrument is given by the following data, taken from reports published by A. van Uden <sup>4)</sup>.

Two groups of deaf-mute children — with approximately the same loss of hearing, viz. more than 100 dB — had to learn by heart six different melodies; the one group had the instrument described here to play on, while the other group learnt

to the music with or without a hearing aid. After an interval of five weeks the two groups were again tested to see if they could still recognize the melodies: the results this time were 86% and 33% respectively. "The motor reactions of the first group were more complete than those of the second: head, trunk, arms and hands were spontaneously moved in time with the music."

Among the deafest children from the third up to the seventh class, who had been playing the instrument for more than six months, 24 were tested on their ability to distinguish between the pitch of different notes. The children were told what pitch interval would be used, and they were to discern when the high and when the low note was played. The outcome was 84% correct answers for the major third, 77% for the minor third, 74% for the major second and 72% for the minor second. The percentage of correct answers for the different ages was: 67% in the third class, 73% in the fourth, 89% in the fifth, 74% in the sixth and 79% in the seventh.

<sup>4)</sup> See the Reports mentioned in footnote <sup>2)</sup>. We are indebted to Mr. Van Uden and Sister Irena for information on the pitch and chord perception tests.



Analogous tests on the acquired ability to distinguish between chords, made on 10 of the deafest children, produced even more remarkable results. In 92% of all cases (12 tests per child) the difference was recognized between a fifth and a second; in 89% between a fifth and a third, in 85% between a fourth and a third, in 81% between a third and a second. Even the difficult differences were recognized between a major second and a minor second (68% of all cases) and between a fifth and a fourth (67% of all cases).

One cannot properly appreciate the problems of teaching the deaf and dumb until one has first sat through a number of lessons in an institute such as that mentioned in this article. After such an experience it is difficult to write objectively on the technical side of the subject without digressing on the human aspect. Those of us with normal hearing have good reason to be thankful for the easy way in which we have been able to learn to speak and to understand music and thereby develop our mental faculties normally.

R. VERMEULEN.

## SMALL BALLASTS FOR FLUORESCENT LAMPS

621.327.534.15.032.434

In general, the designer of fixtures for fluorescent lamps requires that the ballast be as small and light as possible, in order that the fixtures themselves may be economical of materials and attractive in form. However, if the circuit, the type of lamp, the mains voltage and the frequency are fixed, the dimensions of a ballast cannot be reduced — unless special precautions are taken — without this having undesirable consequences: the losses in the ballast and the temperature of the insulation both increase with decrease in dimensions; greater losses mean a lower efficiency and the higher temperature means that for a given insulating material the ballast will have a shorter life.

These deleterious consequences may be quantitatively estimated as follows<sup>1)</sup>. Consider a well-designed ballast and imagine all its linear dimensions to be reduced to  $1/p$  times their original values ( $p > 1$ ): the volume and weight then fall to  $1/p^3$  of their original values. If the power of the lamp is kept constant, then the losses in the copper and iron of the reduced ballast become  $p$  times as great owing to the higher current density and the greater magnetic induction, respectively. This greater loss is dissipated as heat through a  $p^2$  times smaller surface, so that the rise in the temperature of the insulation will be  $p^3$  times greater. If it is required to reduce the volume by, say, 25% ( $p = \sqrt[3]{4/3} = 1.10$ ), then an increase in total loss of 10% and an increase in the rise of temperature of 33% must be expected. The fairly low increase in the loss can sometimes be compensated by more careful design or by using a better magnetic material; the increase in the rise of temperature, however, can only be avoided by ensuring appreciably greater heat removal.

The heat which develops in the coils and cores is transmitted to the walls of the box in which the ballast is mounted, and thence via the fixture to the surroundings. The construction of the fixture and the thermal contact between the box wall and the fixture are highly important in this connection, but will not be discussed in this article since the ballast designer has no direct say in the matter. He can, however, try to make the difference in temperature between the hottest point in the insulation and outside surface as small as possible. For this to be the case, the whole of the space between core-coil assembly and box wall must be filled with a material which is a good conductor of heat. This material,

moreover, must have a number of other properties. In general it is required of the filling material that it should

- a) conduct heat efficiently,
- b) be a good electrical insulator,
- c) damp the acoustic vibrations,
- d) be moisture-repellent,
- e) penetrate well into narrow spaces between the components,
- f) not soften at high temperatures,
- g) have a low thermal expansion, and
- h) be cheap.

We shall now see how far a few of the common filling materials come up to the above requirements.

A material frequently used is asphalt in the form of a casting compound. The coefficient of thermal conductivity of asphalt is about  $10^{-3} \text{ W cm}^{-1} \text{ }^\circ\text{C}^{-1}$ . By admixing quartz sand, ground quartz (which is finer in grain), aluminium oxide or a similar substance, the thermal conductivity can be increased by a factor of 3 to 6. Since the dissipation of choke coils and transformers of ballasts per unit surface area is about  $0.1 \text{ W/cm}^2$ , a temperature gradient of 10 to 5  $^\circ\text{C}$  will occur if the average thickness of the asphalt around the components is 3 mm.

An objectionable property of asphalt, however, is its softening at high temperatures. If the box is completely full, asphalt will be forced out through the seams in the box as a result of thermal expansion. Moreover, the components of the ballast may sink in the soft asphalt, possibly causing short-circuits if they are not attached to the box; and, preferably, in order to avoid hum, they should not be fixed to the box. A way around this is to leave some space in the box and to interpose pieces of insulating paper between the components and the box walls (*fig. 1*), but then the heat transfer suffers appreciably.

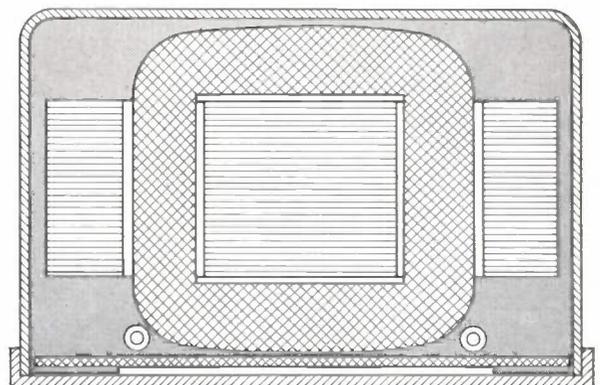


Fig. 1. Cross-section of a ballast filled with asphalt.

<sup>1)</sup> Cf. C. Zwicker, Fluorescent lighting, Philips Technical Library, Eindhoven 1952, p. 128.

The search for a filler which would not soften led some five years ago to the use of aluminium foil. The coefficient of thermal conduction of aluminium is some 300 times greater than that of the best asphalt filler mass. The temperature gradient would therefore be negligibly small if a solid aluminium filling were used. Technically, however, this is not feasible; a large part of the advantage must be sacrificed, since the aluminium has to be in some easily worked form and since the coil has to be enclosed in good electrical insulation. This led to the choice of aluminium foil which has been dimpled in a regular pattern by a previous treatment and is packed into the spaces between the components in the box in packets or in layers (fig. 2). In this way, using only a small

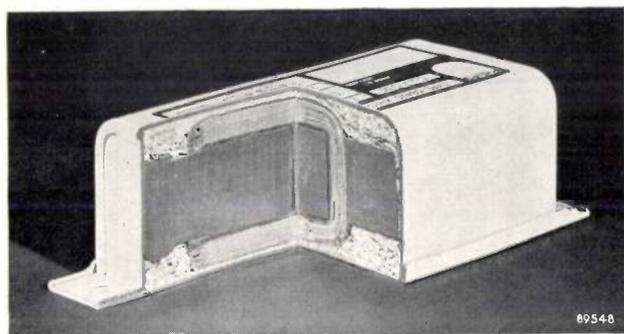


Fig. 2. Section of a ballast with a filling of dimpled aluminium foil.

quantity of aluminium, thermal conduction occurs via a large number of contact points: as a result the heat removal is just as efficient as in the case of complete filling with the best asphalt filler. Moreover, the crumpled aluminium foil is an effective remedy for rattle (the fairly high-pitched sound caused by free mechanical vibrations of choke or transformer cores, as described recently in this journal<sup>2)</sup>).

Although very good results were obtained with aluminium foil, the search for better filling materials has continued, and has been directed notably towards electrical non-conductors which would obviate the need for applying insulation around the coils. Of the electrical non-conductor materials, the polyester group of resins soon attracted attention<sup>3)</sup>. These resins can be cast and are excellent insulators; they are highly moisture-repellent and have a reasonable coefficient of thermal conduction (approx.  $2 \times 10^{-3} \text{ W cm}^{-1} \text{ }^\circ\text{C}^{-1}$ ). After casting, the resin hardens as a result of a chemical reaction and thereafter stays hard, even at the highest temperatures

<sup>2)</sup> E. W. van Heuven, The noise emission of ballasts for fluorescent lamps, Philips tech. Rev. **18**, 110-119, 1956/57, in particular pp. 118-119.

<sup>3)</sup> Miniaturizing with plastics, Brit. Plastics **24**, 302-307, 1951.

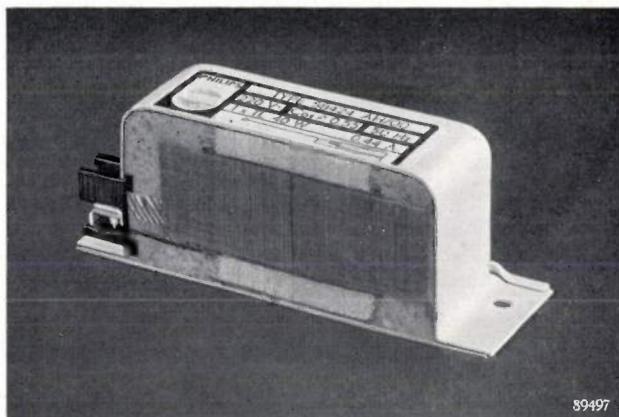


Fig. 3. Section of a ballast with a filling based on polyester resin.

that occur in ballasts. However, it shrinks quite considerably (by about 7%) and its price is high for a filling material. By admixing suitable substances these difficulties can be largely overcome, the shrinkage can be brought down to 1-2%, while at the same time the price is appreciably lowered and the thermal conduction is increased by a factor of 3 to 5. In its unhardened state the viscosity of the mixture can be made such that it penetrates well into narrow gaps; if necessary a gentle pressure can be applied to force it into such gaps. A few spaces will be left nevertheless, but they do not impede heat removal significantly and in practice therefore the ballast box can be considered to be completely filled (fig. 3).

The advantage to be gained with a filling mass based on polyester resin can best be illustrated by means of a few examples.

One of the most important ballasts is the inductive type for the 40 W fluorescent lamp and for 220 V mains. In the existing model with aluminium foil, the box is 122 mm long. In the new model the length has been reduced by 37% to 77 mm, the cross-section remaining the same. The choke coil itself has been reduced in length from 106 mm to 65 mm and has been somewhat enlarged in cross-section,

Table I. Data on 40 W fluorescent lamp ballasts, mains voltage 220 V, old and new forms (with aluminium foil and polyester filler, respectively).

	Aluminium foil	Polyester filler
Total length (mm)	150	105
Length of box (mm)	122	77
Weight (kg)	1.35	0.95
Losses (W)	9	9
Temperature rise ( $^\circ\text{C}$ )*	55	55

\* Measured with a 10% over-voltage and under the test conditions laid down by the International Commission on Rules for the Approval of Electrical Equipment,

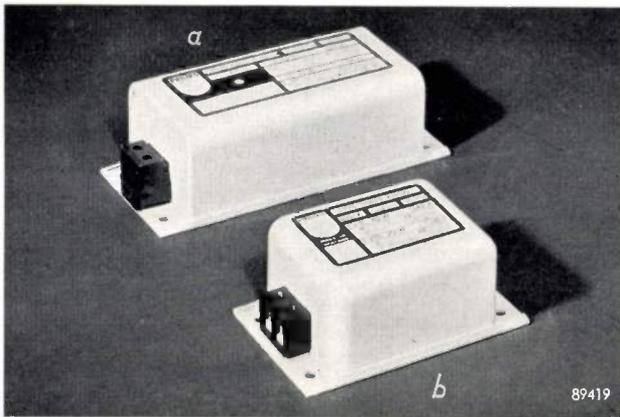


Fig. 4. Ballasts for 40 W fluorescent lamps, mains voltage 220 V. *a* is filled with aluminium foil, *b* with polyester resin.

this having been made possible by the good penetrating power of the filler. *Table I* gives some data on both models, which are shown in *fig. 4*.

The appreciable reduction in length has the additional advantage that the noise produced is considerably less (see the article cited in <sup>2</sup>): firstly, because the elastic movements of the choke or transformer core are smaller, owing to the greater rigidity of the shorter iron circuit, secondly, because the magnetostrictive movements are reduced (for a given magnetic induction they are proportional to the length) and thirdly, because the stray magnetic field is weaker, so that the forces acting on the steel housing are also less. The result of all this is that certain design changes are possible which were not feasible when using the aluminium foil filler, owing to the excessive hum that would be involved.

An example of such design changes is provided by the inductive ballast, again for 40 W fluorescent lamps, but this time for a mains voltage of 110 V. For ignition and stable burning the lamp requires a rated supply voltage of at least 200 V. Such a voltage can be obtained by means of an autotransformer. With the conventional design (i.e. autotransformer with only a very small flux leakage), a separate choke must be incorporated in series with the lamp,

Table II. Data on the ballasts for 40 W fluorescent lamps, mains voltage 110 V, old and new forms.

	Separate transformer and choke; aluminium foil filler	Leakage autotransformer; polyester filler
Total length (mm)	330	195
Length of box (mm)	302	167
Weight (kg)	3.1	1.85
Losses (W)	18	14
Temperature rise (°C) *)	65	55

\*) Measured with a 10% over-voltage and under the test conditions laid down by the International Commission on Rules for the Approval of Electrical Equipment.

just as with a 220 V supply ballast. The conventional ballast thus comprises a transformer of about 100 VA and a choke of about 75 VA, both of which have a low hum level, with a filling of aluminium foil.

A better solution (being smaller and cheaper) is a leakage autotransformer. This is an autotransformer in which special provision is made for flux leakage between the two parts of the coil. A separate choke is then unnecessary if the leakage is given the correct value. *Fig. 5* shows a practical design <sup>4</sup>). Used with

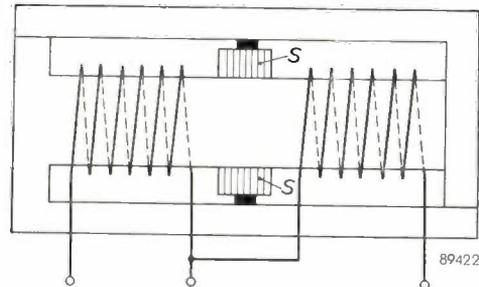


Fig. 5. Leakage autotransformer <sup>4</sup>). The laminated magnetic shunts *S* are driven to the required depth between the two parts of the coil to give the magnetic leakage the correct value; the shunts are then firmly wedged in place.

an aluminium-foil filling, such a leakage transformer with the normal cross-section would have to be about 170 mm long. With a transformer of this length the hum level cannot be suppressed to an acceptable value. Hitherto, therefore, there was no

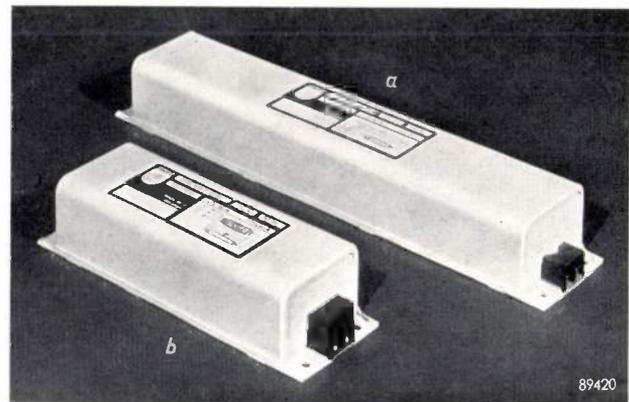


Fig. 6. Ballasts for 40 W fluorescent lamps, mains voltage 110 V. *a*) Separate transformer and choke. *b*) Leakage autotransformer, filling based on polyester resin.

alternative but to use the more costly solution: a separate transformer and choke. With polyester filler it has been possible to reduce the length so as to lower the hum level to a permissible value. Polyester filler therefore has "killed two birds with one stone" in this case. Some data on both models shown in *fig. 6* are given in *Table II*.

Th. HEIENKAMP.

<sup>4</sup>) H. A. W. Klinkhamer, Philips tech. Rev. 1, 342, 1936, fig. 7.

## ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOELAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk \* can be obtained free of charge upon application to the Philips Research Laboratory, Eindhoven, Netherlands.

**2374\***: H. J. G. Meyer: Some aspects of electron-lattice interaction in polar crystals (Dissertation, Amsterdam 1956).

This thesis gives a theoretical treatment of the phenomena which depend on the interaction of lattice vibrations with those electrons which make no contribution to the electrical conductivity. These electrons may be divided into two groups, viz. those which are bound to a defect in a perfect lattice and those which, together with a positive hole in the filled band, are bound to a mobile exciton (Parts I and II respectively of this dissertation). In Part I a survey is given of the effects of lattice vibrations on the transitions of lattice defect electrons — both transitions accompanied by radiation and those which occur without radiation. It is shown — as also in Part II — that all lattice vibrations which interact with electrons have the same frequency. In Part II the effect of the (constant frequency) lattice vibrations on the states of excitons is investigated, on the basis of an "effective mass approximation". The content of this part is substantially that of an earlier publication by the author (see these Abstracts No. 2353).

**2375**: K. H. Klaassens and C. J. Schoot: The reaction between silver phosphate and substituted phenoxyacetyl chlorides (Rec. trav. chim. Pays-Bas 75, 265-270, 1956).

The reaction between silver phosphate and a number of substituted phenoxyacetyl chlorides is described. In all cases liberation of carbon monoxide is observed. After reaction, it is possible in some cases to isolate a substance with the formula  $R-O-CH_2-O-CO-CH_2-O-R$ , where R represents a substituted phenyl group. The remaining compounds form resins of the phenol-formaldehyde type, but under certain circumstances it is possible to identify an *o*-hydroxybenzyl alcohol. A possible mechanism for the reaction is discussed.

**2376**: C. J. Schoot and K. H. Klaassens: Plant growth activity and chemical structure (Rec. trav. chim. Pays-Bas 75, 271-278, 1956).

A hypothesis relating to chemical structure and plant growth activity is formulated. The views of

Hansch and Muir, Smith and Wain and Leaper and Bishop are correlated using as a basis the reaction between  $Ag_3PO_4$  and phenoxyacetyl chlorides described in a previous article. The aryloxyacetic acids, the arylacetic acids and the arylcarboxylic acids are classified in the same scheme. Some indication is given as to the influence of the substituents in the phenoxyacetic acids. The rule of Leaper and Bishop is generalized.

**2377**: J. de Jonge and R. Dijkstra: Some photochemical properties of alkali salts of aryl-diazosulphonic acids (Rec. trav. chim. Pays-Bas 75, 290-300, 1956).

The structure of the labile form of an aryldiazosulphonate and the precise result of exposing a stable diazosulphonate to actinic light is considered to be still uncertain. From experiments on the action of light it can be concluded that a stable diazosulphonate is transformed into a product which is in solution at least partly, but possibly completely dissociated into diazonium ions and sulphite ions. For a number of diazosulphonates the quantum yield for the photochemical reaction has been determined.

**2378**: E. G. Dorgelo: Ueber die Technologie von Magnetrons und Klystrons (Vakuum-Technik 4, 166-176, 1956).

After a short comparison of conventional valves (triodes, tetrodes) with magnetrons and klystrons from the point of view of their application in pulse radar, a description is given of the construction of a few typical valve types. The main part of the article deals with general technological problems in the construction of magnetrons and klystrons. Special attention is paid to the cathode problem. It is the intention of the author to go deeper into the detail questions of manufacture in a second article.

**2379**: J. Bloem and F. A. Kröger: The *p-T-x* phase diagram of the lead-sulphur system (Z. phys. Chemie, Frankfurt a.M. 7, 1-14, 1956).

The phase diagrams (*p-T*, *p-x* and *T-x*) of the Pb-S system, with special reference to the composition of the solid compound PbS, have been deter-

mined by electrical and chemical methods. The maximum melting point is found to lie at 1127 °C and corresponds to a solid PbS phase containing an excess of lead of  $6 \times 10^{18}$  atoms per  $\text{cm}^3$  ( $= 3 \times 10^{-4}$  atoms/mole). Anomalies in the shape of the curves for co-existing solids and liquids indicate the presence of short range order in the liquid at the composition PbS.

**2380:** J. Bloem and F. A. Kröger: The influence of heating in an inert gas atmosphere on the properties of a compound semiconductor, with special reference to PbS (*Z. phys. Chemie, Frankfurt a.M.* 7, 15-26, 1956).

Compounds are stable with respect to heating in a stream of inert gas if the sum of the partial pressures of the components present in the gas phase in equilibrium with the solid has a minimum within the stability range of the solid compound. The solid then attains the composition at which it is in equilibrium with the vapour having the composition of the minimum. The same reasoning applies to heating in vacuo as long as the stationary pressure near the compound does not lie markedly below the minimum equilibrium pressure of the compound. The theory is checked for PbS, pure and doped with silver or bismuth, and good agreement with experiment is observed.

**2381:** J. Haantjes and K. Teer: Compatible colour television (*Wireless Engr.* 33, 3-9 and 39-46, 1956).

In Part 1 a discussion of the requirements and possibilities for a compatible colour-television transmission system is given. A description of a transmission system using two sub-carriers for the additional colour information follows. In Part 2, the two sub-carrier (t.s.c.) system, described in Part 1, is compared with the American N.T.S.C. system. Particular attention is paid to the requirements to be met by the transmission channel, signal-to-noise conditions, cross-talk sub-carrier annoyance and receiver complexity.

**2382:** K. Rodenhuis: Life and reliability of radio tubes for professional equipment (*T. Ned. Radiogenootschap* 21, 65-93, 1956).

Substantially the article which appeared in this Review, vol. 18, pp. 181-192, 1956.

**2383:** G. Diemer and P. Zalm: Voltage distribution inside electroluminescent ZnS crystals (*Physica* 22, 561-562, 1956).

Brief note on a theoretical analysis of the voltage distribution in electroluminescing ZnS crystals,

based on the hypothesis that local internal barriers give rise to high local field strengths.

**2384:** A. van Weel: Some remarks on the radio-frequency phase and amplitude characteristics of television receivers (*J. Brit. Instn. Radio Engrs.* 16, 271-280, 1956.)

The influence on the picture of the steady-state characteristics of the radio-frequency part of a television receiver is considered especially for the frequencies close to the carrier frequency, the so-called Nyquist flank. It follows from numerical calculations that the shape of the amplitude characteristics of this Nyquist flank has but little influence on the picture quality. The performance of a receiver will be substantially the same in combination with a double-sideband transmitter as with a vestigial sideband transmitter, provided the latter has been compensated for its own phase errors. The performance of a vestigial sideband transmitter should be monitored with a phase-linear receiver, of which the exact shape of the Nyquist flank is not very critical.

**2385:** C. J. Heuvelman and A. van Weel: Group-delay measurements (*Wireless Engr.* 33, 107-113, 1956).

A description is given of a simple group-delay meter which, in combination with any conventional wobulator generator, gives the group-delay characteristic directly on an oscilloscope. An automatic gain-control circuit, necessary to maintain a constant level at the output of the network under test, enables the tracing of the amplitude characteristic on a second oscilloscope at the same time. Calibration of amplitude and group-delay scales is possible for any oscilloscope used. A sensitivity of 1 millimicrosecond can be achieved. The frequency range is 20-45 Mc/s.

**2386:** J. S. C. Wessels: De Hill-reactie (*Landbouwkundig Tijdschrift* 68, 647-655, 1956).

Survey article, in Dutch, concerning the Hill reaction; among the points discussed are methods for the study of the Hill reaction, comparison with photosynthesis, the role of the hydrogen-donor and the nature of certain reduction processes.

**2387\*:** F. A. Kröger: Inorganic crystal phosphors (*Ergebn. exakt. Naturwiss.* 29, 61-144, 1956).

Comprehensive survey article on the present state of knowledge of inorganic crystal phosphors. Over 700 references are given.

**2388:** K. ter Haar and J. Bazen: The titration of nickel with EDTA at pH 2.8 (Anal. chim. Acta 14, 409-413, 1956).

A method for determining nickel using EDTA is described. A small excess of EDTA is added, the pH adjusted to 2.7-2.9 and the solution back-titrated with thorium nitrate, against alizarin-S as indicator. Quantities of cobalt up to 10 mg do not interfere. In addition, an approximate determination of the complex-forming constant of Th-EDTA is described, based on the decrease in extinction of an Ni-EDTA solution to which  $\text{Th}^{+4}$  has been added.

**2389:** J. A. Haringx: Non-linearity of corrugated diaphragms (Appl. sci. Res. A6, 45-52, 1956).

Proceeding from the method of calculation for determining the rigidity of corrugated diaphragms given in a former paper, one is able to indicate what degree of non-linearity of the relations between load and deflection is to be expected for large deformations. By means of an example it is shown that the introduction of the corrugations into the flat plate, though unavoidably increasing the initial rigidity, involves an important gain in maximum deflection for the same degree of non-linearity.

**2390:** J. Bloem: p-n junctions in photosensitive PbS layers (Appl. sci. Res. B6, 92-100, 1956).

It has been possible to produce PbS layers containing sharp p-n junctions by partly coating a glass substrate with a trivalent metal and then precipitating a PbS layer from an aqueous solution. A photo-e.m.f. up to 300 mV was observed, corresponding to the value of the energy gap of about 0.3 eV. Some properties of these cells are discussed, attention being paid in particular to the influence of oxygen in the ambient gas.

**2391:** G. W. van Oosterhout: A rapid method for measuring coercive force and other ferromagnetic properties of very small samples (Appl. sci. Res. B6, 101-104, 1956).

A description is given of a method for the quantitative investigation of the permanent magnetic properties, especially the coercive force, of samples of about one milligram. The method is based essentially on the measurement of the alternating e.m.f., generated by letting the sample vibrate in a search coil, with a sensitive voltmeter. For the measurement of magnetic moments the instrument has to be calibrated. It is pointed out how this can be done. The time needed for a measurement is about two minutes.

**2392:** J. J. Opstelten, N. Warmoltz and J. J. Zaalberg van Zelst: A direct-reading double-sided micromanometer (Appl. sci. Res. B6, 129-136, 1956).

On each side of a thin diaphragm, separating two chambers, a condenser plate is mounted at a small distance. The two capacities thus formed are part of a bridge circuit which is fed by a high-frequency oscillator. An amplifier with a narrow bandwidth and a phase-dependent rectifier give a voltage which is fed back to the manometer. At one side of the diaphragm a pressure prevails, much lower than the one to be measured, at the other side the gas is admitted, the pressure of which is to be measured. The displacement of the membrane by the gas pressure is almost completely compensated electrostatically by the feedback voltage. By virtue of a differential way of feedback the reading is proportional to the pressure. The range is from  $10^{-5}$ -0.5 mm of mercury pressure difference at any absolute value. The apparatus is made of chemically fairly resistant materials, and as its indication is independent of the nature of the gas, it can be used for almost every gas. The apparatus can also be used as a non-direct reading instrument, in which case the upper limit of the range is 1 mm of mercury.

Now available:

P. A. Neeteson: Analysis of bistable multivibrator operation. The Eccles-Jordan flip-flop circuit (Philips Technical Library, 82 pp., 34 figures).

The bistable multivibrator, invented by Eccles and Jordan in 1919, attracted little attention for many years, until it suddenly found wide application in pulse techniques. The book reported above (based on a doctoral thesis for the University of Delft) gives for the first time a detailed analysis of the mode of working, particularly of the phenomena which occur when the multivibrator is switching over. A practical result of this analysis has been the development of special valves for circuits of this type.

The twelve chapters of the book bear the following headings: 1. General introduction. 2. Survey of literature. 3. Introduction to the problem. 4. Opening or closing of switches in a network. 5. The static condition of the bistable multivibrator. 6. The dynamic condition. 7. The complete trigger cycle. 8. The trigger sensitivity. 9. The triggering speed. 10. Design considerations. 11. Variations of the fundamental circuit and way of triggering. 12. Conclusion. There follow five appendices with calculations.

# Philips Technical Review

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

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

## FERROXCUBE MATERIAL FOR PIEZOMAGNETIC VIBRATORS

by C. M. van der BURGT.

621.318.134:621.372.412

---

*For generating and detecting ultrasonic waves use is frequently made of piezoelectric or piezomagnetic (magnetostrictive) vibrators. For piezomagnetic vibrator cores, specially prepared types of ceramic ferroxcube have proved in many respects to be as good as, and in some respects even superior to the piezomagnetic metals and alloys that have hitherto commonly been used. Ferroxcube vibrators can be used not only for generating and detecting acoustic waves, but also as temperature-independent elements in electric band-pass filters. Special types of ferroxcube have likewise been developed for this purpose.*

---

### Introduction

In this article we shall be principally concerned with the generation and detection of ultrasonic waves in a surrounding liquid or solid medium. It will hardly be necessary to dwell at length on the great importance that this technique has acquired in recent years. It is used, for example, in under-water signalling (echo-sounding, "Asdic", "Sonar"); for flaw detection in solid materials; for cleaning small mechanical components; for the working of materials, e.g. drilling glass and other brittle materials<sup>1)</sup>; for soldering aluminium (the film of oxide is shattered as it forms); and for therapeutic purposes.<sup>2)</sup>

The ultrasonic vibrators employed in all the above instances belong to the category of transducers which convert electrical energy into mechanical or acoustical energy and vice versa. In this respect they can be compared with loudspeakers and microphones which serve to generate and detect sound waves in air in a frequency range between 20 c/s and 20 kc/s.

In the moving-coil loudspeaker, the conversion of electrical into acoustical energy is based upon the forces, which act on a coil through which the signal current is passed, the coil being located in the field of a permanent magnet. These forces set a membrane (loudspeaker cone) in motion, whereby sound waves are created in the surrounding medium. In principle a loudspeaker can also be made to work as a microphone: when the cone is set vibrating by sound waves, voltages are induced in the coil, i.e. acoustical energy is converted into electrical energy.

Electrodynamical systems cannot in general be used to generate and detect ultrasonic waves (frequencies above 20 kc/s) in liquids or solids; for this purpose vibrators of piezoelectric or piezomagnetic materials are indicated. This article concerns piezomagnetic vibrators, but because of the close analogy with piezoelectric vibrators, it will be useful to begin with a brief discussion of the latter<sup>3)</sup>.

A piezoelectric vibrator may consist, for example, of a quartz crystal which is exposed to an alternating electric field and consequently undergoes periodic alterations in shape. Such a piezo-vibrator is mostly

<sup>1)</sup> An article on ultrasonic drilling will appear in the next issue of this Review.

<sup>2)</sup> See, for example, Technical aspects of sound, Vol. II, editor E. G. Richardson, Elsevier, Amsterdam 1957; T. F. Hueter and R. H. Bolt, Sonics, Wiley & Sons, New York 1955; L. Bergmann, Der Ultraschall, Hirzel Verlag, Zürich, 6th ed. 1954.

<sup>3)</sup> Regarding piezoelectricity, see e.g. J. C. B. Missel, Piezoelectric materials, Philips tech. Rev. II, 145-150, 1949/50.

used at a frequency close to one of the mechanical resonance frequencies of the vibrating body. In the piezoelectric effect of quartz and a number of other crystals, such as ammonium dihydrogen phosphate (ADP), lithium sulphate hydrate (LSH) and potassium sodium tartrate (Rochelle salt), which are used as single crystals, the deformation is an odd function of the electrical field strength and changes in sign with the latter. The deformation of the first three crystals named is proportional to the field strength even up to high values of the latter. (In the case of Rochelle salt which shows marked hysteresis, there is no such simple proportionality.) There are polycrystalline materials, moreover, such as the ceramic materials: barium titanate ( $\text{BaTiO}_3$ ) and lead titanate-zirconate ( $\text{PbTi}_{1-x}\text{Zr}_x\text{O}_3$ ) which are also deformed by an electric field; here, however, deformation is an even function of the field strength, i.e. for a given field strength it is independent of the sign of the field. Such materials, for which of course there is no proportionality between field and deformation, can be rendered "linear" for small alternating fields by imparting them with a biasing polarization. This is done by subjecting them to a constant field on which the low-amplitude alternating field is superimposed. In some cases the polarizing field can be dispensed with, viz. when "remanent electric polarization" can be used.

Let us now compare the above with the piezomagnetic vibrators which can be used for generating and detecting ultrasonic waves<sup>4)</sup>. These vibrators are based on a piezomagnetic effect (magnetostriction) of pre-magnetized (polarized) ferromagnetic materials. This piezomagnetic effect is a linear effect, comparable with the linear piezoelectric effect of polarized barium titanate. In principle all pre-magnetized ferromagnetic materials show this piezomagnetic effect.

To meet the various applications mentioned above, special ferromagnetic materials have been developed whose piezomagnetic effects are particularly pronounced. Initial development work consisted in a search for metals and metal alloys which would fit the purpose, such as nickel, "Permalloy" (Ni, Fe), "Alfer" and "Alfenol" (Al, Fe), and "Vanadium-Permendur" (V, Co, Fe). To generate vibrations, these materials have to be placed in a rapidly alternating magnetic field, and consequently they

must be laminated. If they are not, the eddy currents produced (skin effect) lead to serious weakening of the magnetic field inside the material and thus diminish the resultant piezomagnetic effect.

In recent years ceramic materials have been developed on a wide scale (e.g. the nickel-zinc ferrites, ferroxcube 4), which can be employed as cores in coils, transformers, etc. for use at very high frequencies<sup>5)</sup>. Because of their high electric resistivity these materials need not be laminated. Thus the idea arose of investigating whether such materials could not with advantage be used for ultrasonic piezomagnetic applications.

This was found to be so. The varieties of ferroxcube developed by Philips have been found to be particularly good materials for piezomagnetic vibrators, especially if certain changes are introduced in their chemical composition and method of preparation. The brittleness inherent in ceramics does, it is true, set a limit to the maximum power that can be radiated; but for ferroxcube this limit is high enough for most practical purposes. These new ferroxcube materials can be used both for radiators and for detectors.

Let us return for a moment to the piezoelectric vibrators. Some piezoelectric crystals, such as quartz, possess a high mechanical  $Q$  and hence exhibit strong resonance at a certain sharply defined frequency when an alternating electrical field is applied. This has led to another, non-acoustic application, viz. to their use as the selective components of electrical filters and frequency stabilizers in oscillating circuits. An important feature of such circuits is that the vibration frequency must be independent of temperature. With quartz resonators this is very nearly the case. The constancy can be still further improved by cutting the plate from the crystal in a special manner<sup>6)</sup> (with a prescribed orientation relative to the crystallographic axes).

The question arose as to whether piezomagnetic materials and more particularly ferroxcube could, like piezo-quartz, be applied as elements in electric filters requiring both a high mechanical  $Q$  and a reasonably strong piezomagnetic effect. With ferroxcube, but not with the piezomagnetic metals, this has indeed been found possible. Here again an important factor is the temperature independence of the resonance frequency. Further investigation has shown that the behaviour of ferroxcube

<sup>4)</sup> Piezomagnetism is a new term coined to cover the various reversible magnetomechanical phenomena commonly included under magnetostriction (see for example G. Bradford, *Ultrasonic electro-acoustics: General review, Acustica* 4, 171-181, 1954). A comprehensive historical survey of the application of piezomagnetic vibrators can be found in F. V. Hunt's *Electroacoustics*, Harvard Univ. Press, Cambridge (Mass.) 1954.

<sup>5)</sup> J. J. Went and E. W. Gorter, The magnetic and electrical properties of ferroxcube materials, *Philips tech. Rev.* 13, 181-193, 1951/52.

<sup>6)</sup> W. Parrish, The manufacture of quartz oscillator-plates, *Philips tech. Rev.* 11, 323-332, 351-360, 1949/50, and 12, 166-176, 1950/51.

resonators towards variations in temperature is influenced by the presence of cobalt. It has been found possible with ferroxcubes that do not possess the desired temperature independence to give them the required constancy by adding small amounts of cobalt.

In what follows we shall first briefly discuss piezomagnetism and the working conditions of piezomagnetic vibrators; we shall then go on to examine in greater detail the properties of the new ferroxcube materials for the various applications enumerated above and how those properties are influenced by the manner of preparation and the composition.

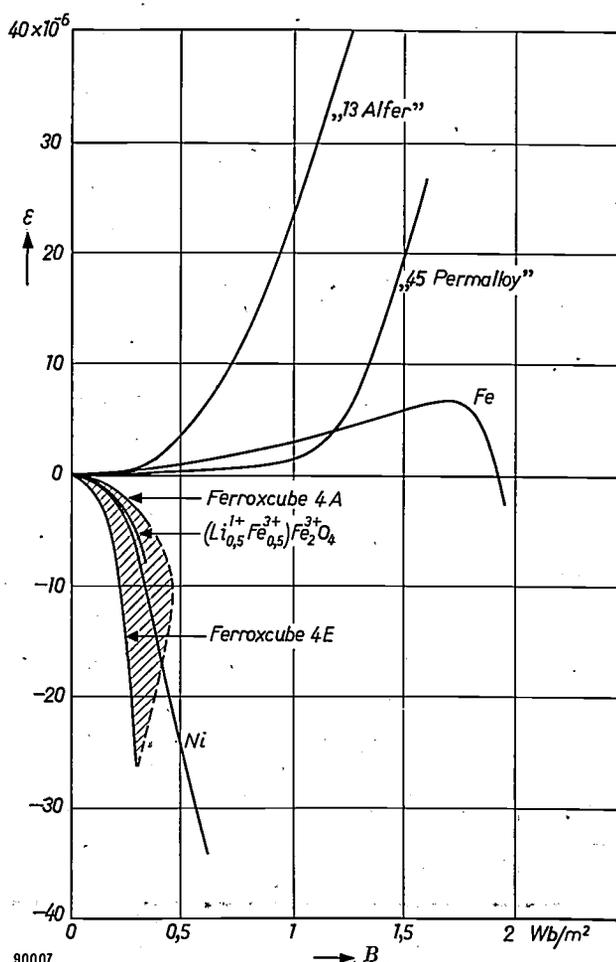
The treatment will be kept as simple as possible. For a more exhaustive analysis the reader is referred to recent publications <sup>7)</sup>, where fuller references to the literature can also be found, and to an article later to appear in this journal which will be devoted to the theoretical aspects of the matter here discussed.

#### Some remarks on piezomagnetism

The magnetism of a ferromagnetic material is caused by the spin vectors of certain groups of electrons. Viewed "microscopically" the material consists of small regions (Weiss domains) in which the spins are more or less parallel below a certain temperature (Curie temperature) and cause a temperature-dependent spontaneous magnetic polarization. If we consider a single crystal which is so small that it contains only one Weiss domain, then a change in the orientation of the spins brought about by an external magnetic field will result in a change in the forces between the atoms. The shape of the crystal will then be changed. If this deformation is opposed by neighbouring crystals, for example when the crystal forms part of a polycrystalline material, internal stresses will be set up in the crystal. If, conversely, the crystal is subjected to external stresses, it will not only be elastically deformed, but moreover, there will be a change in the orientation of the spin vectors and hence a change in the state of magnetization. The relationship between the orientation of the spins, the elastic deformation and the internal and external stresses in a crystal forming part of a polycrystalline material is a complex one. We shall therefore forgo any attempt at a detailed description of the microscopic state of a

piezomagnetic material and shall confine ourselves to the macroscopically perceptible changes.

Let us suppose that we are dealing with a homogeneous polycrystalline, isotropic material. Then the deformation of a rod which is magnetized from the virgin state along its axis, can be described in the usual way by a roughly quadratic relation between the induction <sup>8)</sup>  $B$  and the change in length  $\Delta l$  thereby produced. This relationship which was touched upon in the introduction is shown graphically in *fig. 1* for a number of materials; the tensile strain  $\epsilon = \Delta l/l$  is plotted against  $B$ . A property of all the curves is that they approximately follow the  $B$  axis in the vicinity of the origin. For a number of iron alloys and for ferrous ferrite (magnetite,  $\text{Fe}^{2+}\text{Fe}_2^{3+}\text{O}_4$ ),  $\epsilon$  is positive (positive magnetostriction).  $\epsilon$  is also found to be positive for pure iron, at least for not too high values of  $B$ ; near the saturation point of iron,  $\epsilon$  changes sign.



90007

Fig. 1. Fractional change in length  $\epsilon = \Delta l/l$  for piezomagnetic materials, as a function of  $B$  ( $\approx J$ ). Virgin curves.

<sup>7)</sup> C. M. van der Burgt, Dynamical physical parameters of the magnetostrictive excitation of extensional and torsional vibrations in ferrites, Philips Res. Rep. 8, 91-132, 1953. C. M. van der Burgt, Performance of ceramic ferrite resonators as transducers and filter elements, J. Acoust. Soc. Amer. 28, 1020-1032, 1956.

<sup>8)</sup> It would be more exact to call it the magnetic polarization  $J (= B - \mu_0 H)$ . The difference between  $J$  and  $B$ , however, can in practice be ignored owing to the fairly high (relative) permeability. Moreover, from a technical point of view,  $B$  is the more interesting quantity.

Negative values of  $\epsilon$  are found for pure nickel and for various types of ferroxcube. In the figure the  $\epsilon$ - $B$  curves for ferroxcube 4A ( $(\text{Ni}_{0.35}\text{Zn}_{0.65})\text{Fe}_2\text{O}_4$ ) and ferroxcube 4E ( $\text{NiFe}_2\text{O}_4$ ) are drawn. The curves for the other nickel-zinc ferrites (ferroxcube 4B, 4C, 4D) lie between them in the shaded portion. Also drawn in this shaded part is the curve for lithium ferrite ( $(\text{Li}_{0.5}^{1+}\text{Fe}_{0.5}^{3+})\text{Fe}_2^{3+}\text{O}_4$ ). All the curves, except that for iron, are continued to near the saturation point. For the ferroxcube 4 materials the saturation points lie on the dotted boundary line of the shaded portion. The value of  $\epsilon$  at saturation, i.e. the saturation magnetostriction  $\epsilon_s$ , can be found in Table I and is a measure of the suitability of a material for high-power piezomagnetic radiators. For metals and alloys  $\epsilon_s$  varies from  $70 \times 10^{-6}$  ("2V-Permendur") to  $-33 \times 10^{-6}$  (Ni), and for ferrites from  $40 \times 10^{-6}$  (ferrous ferrite) to  $-27 \times 10^{-6}$  (nickel ferrite). The permanently magnetic materials cobalt and cobalt ferrite are not taken into account here (although  $\text{CoFe}_2\text{O}_4$  does appear in the table).

Additional data given in the table, besides  $\epsilon_s$  and the saturation polarization  $J_s$  ( $\approx$  saturation induction  $B_s$ ), are: the D.C. electrical resistivity  $\rho_{\text{el}}$ , the Curie temperature  $T_c$ , and the coupling coefficient at remanence  $k_{\text{rem}}$ . A high Curie temperature is of interest if the material is to be used at high temperatures. Disregarding "Permendur" which has a very high  $T_c$  and ferroxcube 4A which has a very low  $T_c$ , it will be seen that in this respect the ferrites are comparable with alloys. We shall come back to the significance of the coupling coefficient later.

In fig. 1 no account is taken of hysteresis. The figure holds for materials which are in the virgin

state when the (macroscopic) induction is zero. In fig. 2a the general relationship between  $B$  and  $H$  is represented schematically; both the virgin curve and the hysteresis loop are shown. In fig. 2b the corresponding relationship between  $\epsilon$  and  $H$  is given. The

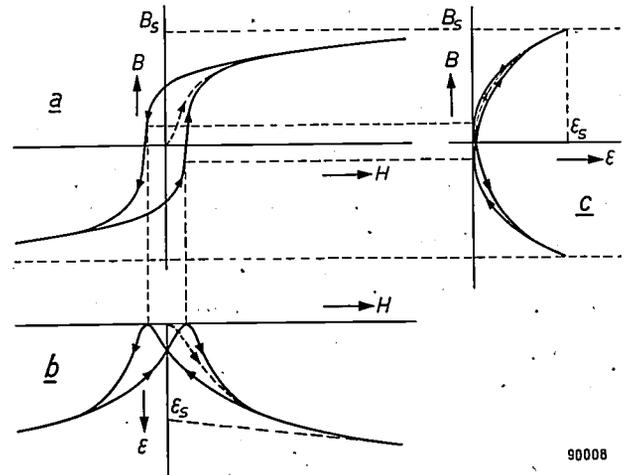


Fig. 2. a)  $B$  as a function of  $H$ ; b)  $\epsilon$  as a function of  $H$ ; c)  $\epsilon$  as a function of  $B$ . Both the hysteresis loops and the virgin curves are drawn (static behaviour).

sign of the magnetostriction is independent of the sign of  $H$ , hence a butterfly-shaped curve is obtained. Fig. 2c shows the roughly quadratic relationship between  $\epsilon$  and  $B$ .

If, starting from a point on the virgin curve,  $H$  is allowed to decrease slightly and then to increase (see fig. 3), a narrow loop is described in the  $B$ - $H$  diagram. The slope of this loop corresponds to the "reversible permeability". A similar loop is described in the  $\epsilon$ - $H$  diagram. It is assumed that there is no

Table I. Properties of various materials for piezomagnetic vibrators. For the ferrites the molecular composition is indicated; for the alloys the composition is quoted in percentages by weight. The porosity  $p$ , on which the saturation polarization  $J_s$  of the ferrites depends, is defined as pore volume/total volume.

Material	Composition	$10^6 \epsilon_s$	$J_s$ (Wb/m <sup>2</sup> )	$\rho_{\text{el}}$ ( $\Omega\text{m}$ )	$T_c$ ( $^{\circ}\text{C}$ )	$ k_{\text{rem}} $
"2V-Permendur"	2 V, 49 Co, 49 Fe	+ 70	2.4	$2.5-4 \times 10^{-7}$	980	0.19-0.26*
"13 Alfer"	13 Al, 87 Fe	+ 40	1.3	$9 \times 10^{-7}$	$\sim 500$	0.19-0.26*
"45 Permalloy"	45 Ni, 55 Fe	+ 27	1.6	$5-7 \times 10^{-7}$	$\sim 440$	0.11-0.17*
Nickel	99.9 Ni	- 33	0.63	$0.7 \times 10^{-7}$	358	0.14-0.23*)†
Nickel-cobalt alloy	96 Ni, 4 Co	- 31	0.68	$1.0 \times 10^{-7}$	410	0.34*)†
Ferrous ferrite	$\text{Fe}^{2+}\text{Fe}_2^{3+}\text{O}_4$	+ 40	0.61 (1- $p$ )	$\sim 10^{-4}$	580	—
Ferroxcube 4A	$\text{Ni}_{0.35}^{2+}\text{Zn}_{0.65}^{2+}\text{Fe}_2^{3+}\text{O}_4$	- 5	0.40 (1- $p$ )	$\geq 10^4$	190	0.06-0.10**
Lithium ferrite	$(\text{Li}_{0.5}^{1+}\text{Fe}_{0.5}^{3+})\text{Fe}_2^{3+}\text{O}_4$	- 8	0.39 (1- $p$ )	$\geq 10^4$	640	0.03-0.05**
Ferroxcube 4E	$\text{Ni}^{2+}\text{Fe}_2^{3+}\text{O}_4$	- 27	0.33 (1- $p$ )	$\geq 10^4$	590	0.14-0.20**
Nickel-cobalt ferrite	$\text{Ni}_{0.95}^{2+}\text{Co}_{0.02}^{2+}\text{Fe}_2^{3+}\text{O}_4$	- 26	0.33 (1- $p$ )	$\geq 10^4$	590	0.22-0.25**
Cobalt ferrite	$\text{Co}^{2+}\text{Fe}_2^{3+}\text{O}_4$	$\sim -200$	0.54 (1- $p$ )	$\geq 10^4$	510	—

\*) Dependent on the mechanical and thermal treatment of the laminations.

\*\*\*) Dependent on the porosity, i.e. the sintering temperature. The high value reported is that found for a certain optimum porosity ( $p$ , in Table II). The low values reported for ferroxcube 4A and 4E are those of the currently available commercial materials, for which the porosity  $p$  is given in Table II.

†) C. A. Clark, The dynamic magnetostriction of nickel-cobalt alloys, Brit. J. appl. Phys. 7, 355-360, 1956.

external tensile stress  $\sigma$  acting on the rod during these variations in  $B$  and  $\epsilon$ . The slopes of the loops are therefore designated by  $(\partial B/\partial H)_\sigma = \mu^\sigma$  and  $(\partial \epsilon/\partial H)_\sigma = d$  respectively <sup>9)</sup>. It is also possible to

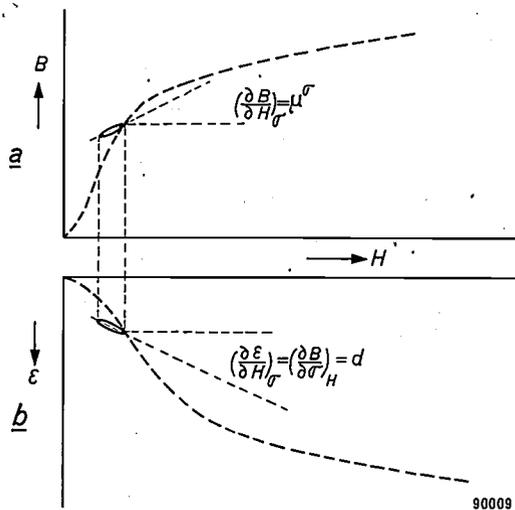


Fig. 3. Dynamic behaviour of  $B$  and  $\epsilon$  as functions of  $H$ , for normal piezomagnetic material.

vary  $H$  and at the same time to subject the rod, for example by clamping it, to such forces that no change in length  $\epsilon$  can occur. A different value is then obtained for the permeability which we designate by  $(\partial B/\partial H)_\epsilon = \mu^\epsilon$ . Piezomagnetic behaviour manifests itself in the fact that  $\mu^\sigma$  and  $\mu^\epsilon$  have different values, i.e.  $\mu^\sigma > \mu^\epsilon$ . If, starting from a state in which  $B = B_0, H = H_0$ , we subject the rod to an alternating magnetic field with amplitude  $\hat{H}$ , then for the unclamped rod the maximum magnetic energy is  $\frac{1}{2}\mu^\sigma \hat{H}^2$  and for the clamped rod only  $\frac{1}{2}\mu^\epsilon \hat{H}^2$ . In the first case, however, internal elastic energy is supplied to a value of  $\frac{1}{2}(\mu^\sigma - \mu^\epsilon)\hat{H}^2$ . As can be seen from the above, the ratio

$$k^2 = \frac{\mu^\sigma - \mu^\epsilon}{\mu^\sigma} \dots \dots \dots (1)$$

gives the fraction of the magnetic energy which can be converted into mechanical energy. The root of this quantity, i.e.  $k$ , is the magnetomechanical or piezomagnetic coupling coefficient reported in Table I. For ferrites  $k^2$  may have a value up to 0.10.

The quantity  $k^2$  must be clearly distinguished from the electroacoustic efficiency  $\eta_{ea}$  to be discussed later (the latter is the percentage of the electrical power supplied that is converted into acoustic power) which can attain values of almost 100%.

<sup>9)</sup> Following the standardized system adopted in the U.S.A. for the analogous piezoelectric case (Proc. Inst. Radio Engrs. 37, 1378-1395, 1949), we print the variable which is kept equal to zero or, in the general case, is kept constant, as a superscript and not as a subscript of the quantities  $\mu$  and  $E$  (permeability and modulus of elasticity).

The rod can, of course, also be subjected to an alternating tensile stress without being subjected to a magnetic field. The specific elastic compliance can be determined in this way. Since the determination is carried out with  $H$  constant, the compliance found is designated by  $(\partial \epsilon/\partial \sigma)_H = 1/E^H$ . If the rod thus strained is in the virgin state ( $H_0 = 0, B_0 = 0$ ), no magnetic changes are perceptible. The macroscopic induction is and remains zero. If, however, the rod is initially in a magnetized state ( $B_0 \neq 0$ ), the extension results in a change of induction. The quantity in question  $(\partial B/\partial \sigma)_H$  is found to equal the quantity  $(\partial \epsilon/\partial H)_\sigma$  discussed above, so that  $(\partial B/\partial \sigma)_H$  is likewise equal to  $d$ . If it is required to suppress any change in induction brought about by the strain, the rod must be surrounded by a short-circuited solenoid having zero resistance. Currents will then be set up in the turns of this coil, so as to keep  $B$  constant. The resulting compliance  $(\partial \epsilon/\partial \sigma)_B = 1/E^B$  differs from the compliance  $1/E^H$  just discussed, and in fact  $1/E^H > 1/E^B$ . It is also found that

$$k^2 = \frac{1/E^H - 1/E^B}{1/E^H} \dots \dots \dots (2)$$

Because magnetic and elastic losses occur with alternating magnetic and mechanical loading, the

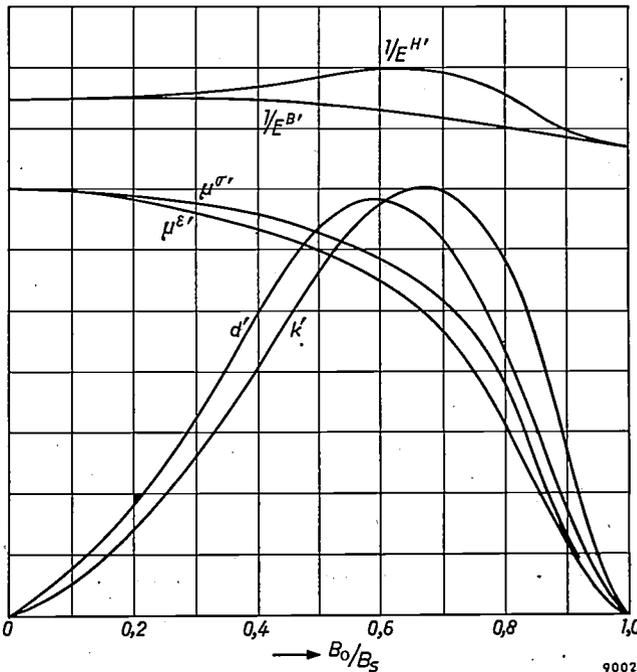


Fig. 4. Schematic curves for  $\mu^\sigma, \mu^\epsilon, 1/E^H$  and  $1/E^B$ , together with the sensitivity constant  $d'$  and the coupling coefficient  $k'$ , as functions of  $B_0/B_s$ , for a ferroxcube. Virgin curves, i.e. the curves for the reversible dynamic constants shown hold for working points on the virgin curve of increasing magnetic bias.

quantities  $\mu^\sigma$ ,  $\mu^s$ ,  $1/E^H$ , and  $1/E^B$  are complex<sup>10</sup>. In fig. 4 the real parts  $\mu^{\sigma'}$ ,  $\mu^{s'}$ ,  $1/E^{H'}$  and  $1/E^{B'}$  of these quantities as well as the real parts  $d'$  and  $k'$  of  $d$  and  $k$  are plotted as functions of  $B_0/B_s$ .

**The behaviour of piezomagnetic vibrators in the resonance band**

The two main forms of piezomagnetic vibrators are the rod vibrator and the ring vibrator, consisting respectively of a rod or ring-shaped core made of a piezomagnetic material and surrounded by a coil. Like any other solid body the core can perform free elastic vibrations having a number of modes of vibration and corresponding characteristic frequencies. We shall confine ourselves to the fundamental frequency. The fundamental frequency of the rod on open circuit (i.e. in the absence of an alternating magnetic field) corresponds to a longitudinal vibration with a displacement node (strain antinode) at the centre and displacement antinodes (strain nodes) at either end. The length  $l$  of the rod corresponds to half a wavelength. Since the product of frequency and wavelength is equal to the velocity of sound  $v^H = (E^H/\rho)^{1/2}$  in the material, the frequency of free vibration is given by

$$f^H = \frac{1}{2l} \left( \frac{E^H}{\rho} \right)^{1/2}, \dots \dots (3a)$$

where  $\rho$  is the specific gravity.

In the case of the ring the fundamental frequency corresponds to a mode of vibration in which the circular shape is preserved, but the diameter  $2R$  is a periodic function of the time. For the ring we have

$$f^H = \frac{1}{2\pi R} \left( \frac{E^H}{\rho} \right)^{1/2} \dots \dots (3b)$$

For formula (3a) we must not use for  $E^H$  the value discussed above, which we find for homogeneous strain, but an average value which is somewhat higher than  $E^H$ . The reason for this is connected with the fact that the alternating induction along a half-wave vibrator may be practically constant, whereas the alternating strain is a sinusoidal function of position, as in the window-type half-wave vibrator to be discussed in the next section.

The rod or ring can also be allowed to vibrate elastically with the coil short-circuited. This being the case,  $B$  is constant and  $v^B = (E^B/\rho)^{1/2}$ . The corresponding vibration frequency is

$$f^B = \frac{1}{2l} \left( \frac{E^B}{\rho} \right)^{1/2} \dots \dots (4a)$$

<sup>10</sup> They are therefore of the form  $a = a' - ja'' = a'(1 - j\delta)$ , where  $a''$  is always much smaller than  $a'$  and consequently the loss angle  $\delta \ll 1$ . Further  $|a| = \sqrt{a'^2 + a''^2} \approx a'$ . The mechanical  $Q$  and magnetic  $Q$  later to be discussed are defined by  $Q = 1/\delta$ .

for the rod, and

$$f^B = \frac{1}{2\pi R} \left( \frac{E^B}{\rho} \right)^{1/2} \dots \dots (4b)$$

for the ring. The frequencies  $f^H$  and  $f^B$  lie close together, and

$$\frac{f^B - f^H}{f^B} \approx \frac{1}{2} |k^2|/\varphi, \dots \dots (5)$$

where  $\varphi \approx \pi^2/8$  for the rod and  $\varphi = 1$  for the ring.

If an alternating current is passed through the coil, then owing to the piezomagnetic effect, the length of the polarized rod or the diameter of the polarized ring (as the case may be), will be a periodic function of the time. The rod or the ring then performs forced vibrations whose amplitude in the resonance band may be large compared with their amplitude outside this band, as a result of the high mechanical  $Q$ .

The behaviour in the resonance band can be investigated by measuring the impedance  $Z$  or the admittance  $Y (= Z^{-1})$  as a function of the frequency. In fig. 5 the absolute values  $|Z|$  and  $|Y|$  are plotted as functions of  $f$ . It will be seen that  $|Z|$  has a

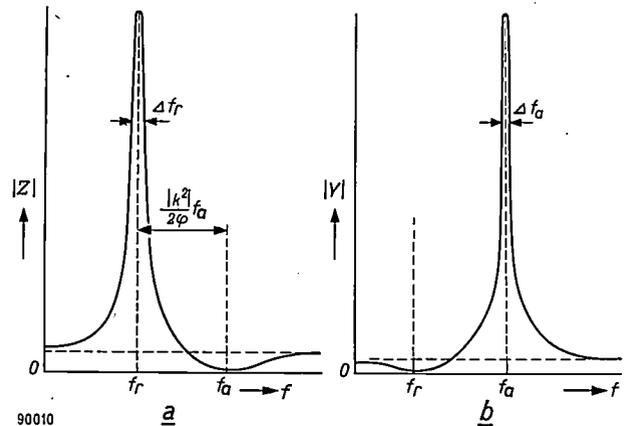


Fig. 5. Absolute values for the impedance  $Z$  and the admittance  $Y$  of a ferrocube vibrator (in air), as functions of the frequency  $f$ . Here  $f_a - f_r \approx f^B - f^H \approx \frac{1}{2} (|k|^2/\varphi) f^B$ , where  $\varphi = 1$  for ring vibrators and  $\varphi \approx \pi^2/8$  for half-wave vibrators. At anti-resonance the mechanical  $Q (= f/\Delta f)$  is considerably higher than at resonance.

maximum at frequency  $f_r$  ("resonance frequency") and a minimum at frequency  $f_a$  ("anti-resonance frequency"). For  $|Y|$ , of course, the maximum and minimum are interchanged. If the maximum value of  $|Z|$  is many times greater than the value of  $|Z|$  outside resonance (say at least 10 times), then to a very close approximation  $f_r = f^H$  and  $f_a = f^B$ . Consequently between  $f_r$  and  $f_a$  there likewise exists the relationship

$$\frac{f_a - f_r}{f_a} \approx \frac{1}{2} |k^2|/\varphi, \dots \dots (6)$$

The above refers to a non-radiating vibrator, i.e. when it is vibrating in air (the radiation is then negligible) or in vacuum, and when the coil is so mounted that the core can vibrate without touching it. If the vibrator is in contact with a liquid, acoustic waves are launched into the latter, and the vibrator is strongly damped. Intimate contact between vibrator core and coil and in the case of a laminated core, friction between the laminations can also contribute to the damping.

Since a liquid in contact with the radiating surface of the vibrator will move with the latter, the apparent mass of the vibrator will increase, at least, if no plane acoustic waves are launched. As a result the resonance frequencies will be shifted, becoming somewhat lower. Simultaneously, damping will flatten and broaden the maxima of  $|Z|$  and  $|Y|$ .

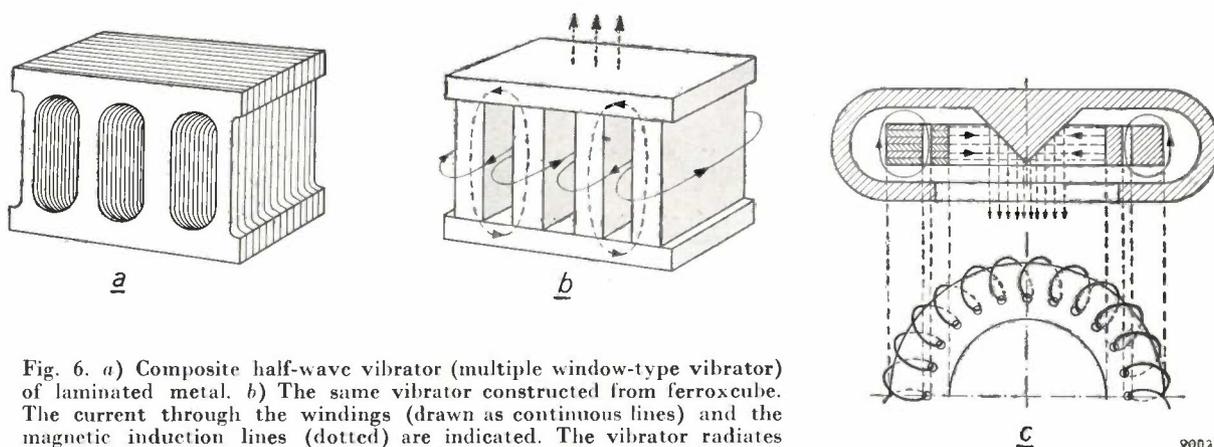


Fig. 6. *a*) Composite half-wave vibrator (multiple window-type vibrator) of laminated metal. *b*) The same vibrator constructed from ferroxcube. The current through the windings (drawn as continuous lines) and the magnetic induction lines (dotted) are indicated. The vibrator radiates from the top surface (dotted-line arrows). *c*) Ring vibrator with conical reflector; to show how the metal construction would appear, lamination of the ring is drawn on the left of the centre line. The direction of radiation is indicated by dotted-line arrows. Radiation from the surface opposite and parallel to the radiating surface is prevented by a sheet of acoustic shielding material such as cell-tight foam rubber.

Damping leads moreover to the dissipation of power of which only the radiated power is useful. The other types of dissipation (magnetic and elastic) must be regarded as losses, and give rise to heating of the vibrator core and coil.

#### Some conventional forms of piezomagnetic transducers

The ring and rod vibrators discussed above are actually employed in practice. Rod vibrators are often in the form of several rods or plates arranged alongside each other and connected together by means of end plates (window-type vibrators). In *fig. 6* piezomagnetic transducers such as are used at frequencies of about 30 kc/s, are shown schematically.

*Fig. 6a* represents a half-wave radiator of piezomagnetic metal, showing the laminations. The latter

are about 0.1-0.2 mm thick for nickel and about 0.2-0.4 mm thick for "Alfer" and "Permendur".

*Fig. 6b* represents a corresponding ferroxcube vibrator. The core is not laminated. The legs are of strongly piezomagnetic ferroxcube (e.g. nickel-cobalt ferrite). The end plates are made of non-piezomagnetic or weakly piezomagnetic material having a high saturation polarization (e.g. manganese-zinc ferrite or nickel-zinc ferrite). To ensure linearity the core is given a biasing polarization as described in the introduction. This polarization (which may be remanent polarization) is parallel to that of the alternating induction (indicated in the figure).

To avoid the necessity of a strong direct current for the biasing of high-intensity radiators, one may instead cement in each magnetic circuit one or two thin slices of permanent-magnet ceramic material<sup>11)</sup>, for example ferroxdure<sup>12)</sup>.

In *fig. 6c* a "breathing" ring vibrator with reflector is shown (constructed in laminated metal or ferroxcube). The turns are led through holes in the ring so that the radiating surface (here the inside surface) is left free. Ring vibrators also exist in which radiation takes place from the outside surface.

*Fig. 7* is a photograph of a ferroxcube half-wave transducer with and without windings.

#### The generation of acoustic waves

The most important quantities involved in the generation of acoustic waves in a medium are: the

<sup>11)</sup> Y. Kikuchi and co-workers, Magnetostrictive ultrasonic transducers made of ferrites, *Sci. Rep. Res. Insts. Tôhoku Univ. (B)* 7, 9-15, 1955.

<sup>12)</sup> J. J. Went, G. W. Rathenau, E. W. Gorter and G. W. van Oosterhout, Ferroxdure, a class of new permanent magnet materials, *Philips tech. Rev.* 13, 194-208, 1951/52. See also an article by H. G. Bruijning and A. Rademakers, to appear shortly in this Review.

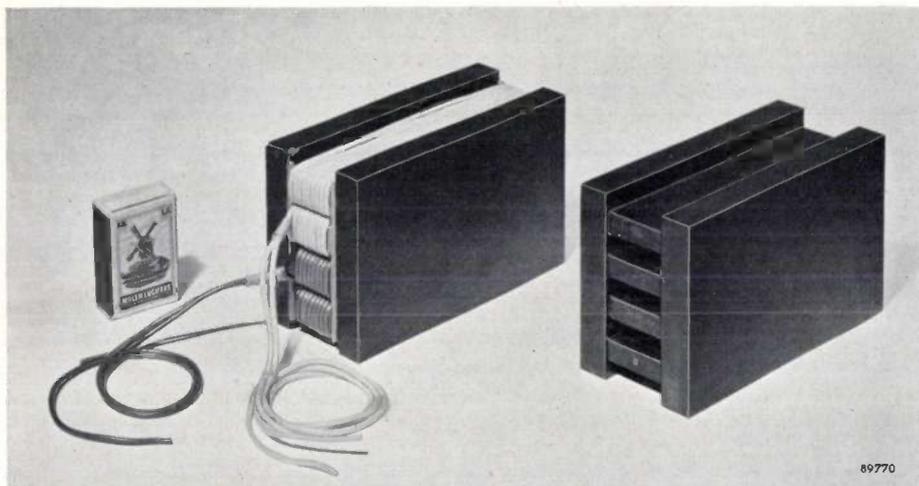


Fig. 7. A photograph of a ferrocube half-wave transducer (multiple window type), with windings and without.

electroacoustic efficiency  $\eta_{ea}$  and the maximum acoustic power radiated.

The energy consumed is the sum of all the dissipations. The electroacoustic efficiency is therefore:

$$\eta_{ea} = \frac{\text{dissipation by radiation}}{\text{sum of all dissipations}} \quad (7)$$

A high efficiency is desirable in ferrocube vibrators, not so much because of power consumption as because of the fact that dissipation of energy in the vibrator leads to heating of the vibrator core and hence to thermal stresses. Owing to the brittleness of the material this would lower the maximum permissible alternating strain and consequently the maximum safe intensity. We shall return to this point a little later.

*The maximum electroacoustic efficiency*

For a given vibrator the values of the losses and of the dissipation by radiation, hence also of  $\eta_{ea}$ , depend upon the frequency of the alternating magnetic field.

In fig. 8b the electric power supplied, at almost constant voltage ( $\hat{V} \propto f^{1/2}$ ), is plotted as a function of the frequency. As fig. 8a shows, this power can be derived from the complex diagram for the quantity  $fY$  ( $Y =$  the admittance of the vibrator), which is proportional to  $1/\mu$  (where  $\mu$  is the complex permeability of the vibrator). In fig. 8a and b the state of affairs is shown with the vibrator in air and in water.

It is now found that, if the copper losses can be ignored, the highest value for  $\eta_{ea}$  is obtained when the vibrator is operated at the frequency  $f_{max}$  ( $\approx f^B \approx f_a$ ) at which the real part  $Re(fY)$  is a

maximum<sup>13) 14)</sup>. In what follows we shall assume that this is always the case. We then have<sup>15)</sup>:

$$\eta_{ea \max} = \frac{D_w}{h + D_w} \frac{D_v - D_w}{D_v} \quad (8)$$

We shall not go into the derivation of this formula here. (This will be discussed in a second article to appear in this Review.)  $D_v$  and  $D_w$  are the diameters of the admittance circles (fig. 8a) when the completely assembled vibrator is set up in air (or in vacuum) and in water respectively;  $h$  is the smallest distance from the admittance circles to the imaginary axis (see fig. 8b).  $D_v$  and  $D_w$  are proportional to  $|k^2|$ ; further  $D_v$  is proportional to the mechanical  $Q$  in air (see<sup>10)</sup>), and  $h$  is inversely proportional to the

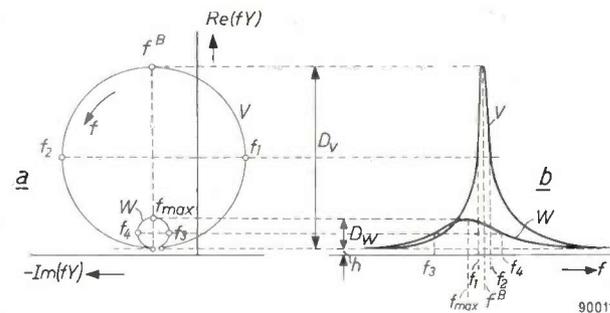


Fig. 8. a) Complex diagram of  $fY$  with  $f$  as parameter, for a vibrator in air or vacuum ( $V$ ) and in water ( $W$ ). In reality the ratio of the diameters of the two circles is between 10 and 100. b) The real part  $Re(fY)$  as a function of  $f$ . For an approximately constant voltage ( $\hat{V} \propto f^{1/2}$ ),  $Re(fY)$  is proportional to the electrical power supplied to the vibrator.

magnetic  $Q$ . Hence,  $\eta_{ea \max}$  increases with increasing mechanical and magnetic  $Q$ 's and with increasing  $k$ .

If we compare vibrators of different dimensions, whose working frequencies ( $= f_{max}$ ) will therefore also be different, then  $\eta_{ea \max}$  can be plotted as a

<sup>13)</sup> Y. Kikuchi and H. Shimizu, On the "B-type" resonance of magnetostriction and electrostriction transducers, Sci. Rep. Res. Insts. Tôhoku Univ. (B) 4, 173-203, 1952.  
<sup>14)</sup> H. Thiede, Untersuchungen an Ferriten auf ihre Eignung als Flüssigkeitsschallwandler, Acustica 4, 532-536, 1954; see also page 122 et seq. of the book by Hunt cited in note<sup>4)</sup>.  
<sup>15)</sup> Design and construction of magnetostriction transducers, Summary Technical Report, Div. 6, NDRC, Vol. 13, Washington (D.C.) 1946, page 59 (PB Report No. 77 669).

function of this frequency. In addition  $\eta_{ea \max}$  will depend upon the shape of the vibrator and on the nature of the piezomagnetic material, and also, to a small extent, on the power radiated. In fig. 9,  $\eta_{ea \max}$

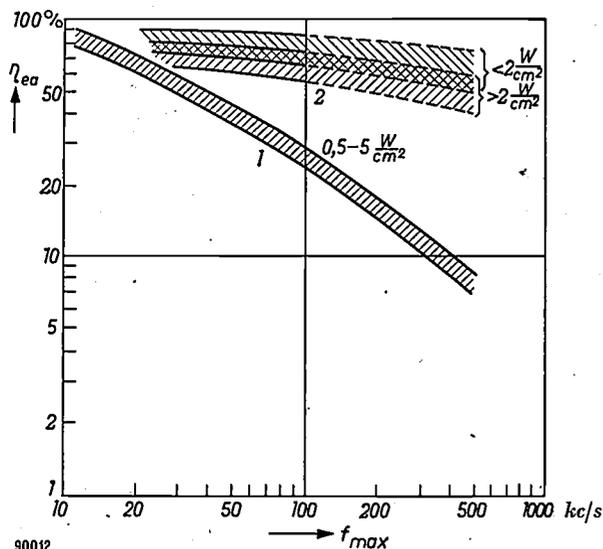


Fig. 9. Electroacoustic efficiency  $\eta_{ea}$  for vibrators in continuous operation (working constantly at their optimum frequency  $f_{max}$ ) as a function of  $f_{max}$ ; curves 1 for metals, 0.5-5 W/cm<sup>2</sup> (data from the literature), curves 2 for ferroxcube, <2 W/cm<sup>2</sup> and >2 W/cm<sup>2</sup> (very few measurements have been made above 100 kc/s).

is plotted roughly as a function of  $f_{max}$  for window-type radiators. The curves given refer to laminated metallic materials, at a radiation intensity of 0.5 to 5 W/cm<sup>2</sup>, and to suitable types of ferroxcube, at intensities < 2 W/cm<sup>2</sup> and > 2 W/cm<sup>2</sup>. It can be seen that even for such low frequencies as 20 kc/s, ferroxcube is preferable to metal. The superiority of ferroxcube is due to the absence of eddy-current losses and frictional losses between laminations.

*The maximum safe load of ferroxcube radiators*

A second important point in the generation of acoustic waves is the maximum safe load. This is determined by the mechanical strength of the material, for during vibration it is subjected in turn to tension and compression. Now, it is well known that the ultimate unidirectional compressive stress of ceramic materials exceeds their ultimate tensile stress by a factor of 10-30, so that the tensile strength is the determining quantity. An idea of the magnitude of the tensile strength can be obtained from easily performed bending tests. The tensile strength, the bending strength and the compressive strength all fall as the porosity  $p$  (pore volume/total volume) increases. This can be seen from fig. 10 in which the compressive strength and the tensile strength (the latter from bending and tensile tests) of ferroxcube 4

materials are plotted as functions of  $p$ . It will be noted that the tensile strength calculated from bending tests is somewhat too high for dense ferrites and somewhat too low for porous ferrites, compared with the tensile strength found in very carefully performed tensile tests. When using these data it must also be taken into account that the tensile strength of ceramic materials is a function of the cross-sectional area — or better, of the volume — of the material, since the fracture in ceramic materials is essentially a matter of statistics. Fracture is initiated at the weakest point in the volume under load: the probability of there being a weak point increases as the volume increases. There are various indications that the tensile strength, as reported in fig. 10, should be divided by a factor of 2 to 3 for very large cross-sections in order to obtain the correct value. The experimental curves given in fig. 10 refer to cross-sections of 2 mm × 3 mm.

It has been found in practice<sup>16)</sup> that even rods of ferroxcube 4D (length 6 cm, cross-section 2 cm<sup>2</sup>), which has not been specially prepared for piezomagnetic applications and having (in this respect) an unfavourably high porosity ( $p = 0.20$ ), can continuously deliver an intensity of 8 W/cm<sup>2</sup> at 40 kc/s from both ends with an efficiency greater than 0.7.

It can be safely assumed that if the material is judiciously chosen (good homogeneity, small  $p$  and large  $k$ , and hence a large  $\eta_{ea \max}$ ), it will be possible to obtain continuous radiation intensities of at least 10 W/cm<sup>2</sup> at 50 kc/s without any great risk of fracture. Systematic tests on this score are still in progress. It can be taken as a rule of thumb that the

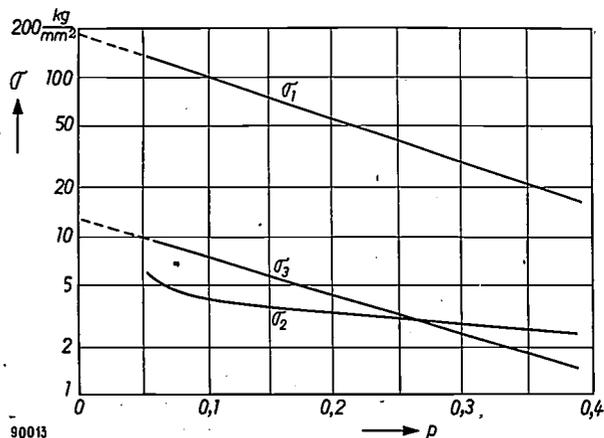


Fig. 10. The compressive strength ( $\sigma_1$ ) and the tensile strength from tensile tests ( $\sigma_2$ ) and bending tests ( $\sigma_3$ ) as functions of the porosity  $p$ . The curves are roughly the same for the various types of ferroxcube. All refer to a cross-section of 2 mm × 3 mm.

<sup>16)</sup> U. Enz, Erzeugung von Ultraschall mit Ferriten, Dissertation E.T.H. Zürich 1955 (page 45).

maximum safe intensity for continuous loading (C.W. operation) is  $15(1 - 2p)$  W/cm<sup>2</sup>.

The figures quoted for the radiation intensity in W/cm<sup>2</sup> refer to simple half-wave bar-type radiators. The area involved is then simply the cross-sectional area of the bar. For window-type radiators the values quoted should be divided by roughly 1.5 (for single window types) or by 1.5 - 2 (for multiple window types), in order to find the values per square centimetre of the radiating surface.

Intensities of that magnitude are perfectly satisfactory for many practical applications. If instead of continuous loading, the vibrator is pulse-operated (whereby temperature gradients and hence thermal stresses in the material are avoided), ferroxcube radiators can be expected to give safe radiation intensities of 15 to 20 W/cm<sup>2</sup> in water.

*Types of ferroxcube suitable for ultrasonic radiators*

We have already seen that the porosity of the ceramic radiators constitutes a limiting factor with respect to the maximum safe intensity. It will however be clear that one should first ensure that the amplitude of the piezomagnetic deformation of the material for reasonable alternating magnetic fields is such that the fracture limit is indeed the determining factor. In practice this means that a strain amplitude  $\hat{\epsilon} = 10^{-4}$  must occur at resonance. If we suppose that outside resonance a strain amplitude of up to about one third of the saturation magnetostriction  $\epsilon_s$  may be produced in the pre-magnetized vibrator by a reasonable field strength, we find that on account of the mechanical  $Q$  (in water) of 15-45,  $|\epsilon_s|$  must be equal to or greater than  $10 \times 10^{-6}$ .

Ferroxcube materials of the types thus far available on the market are prepared with special regard to optimum properties for specific magnetic applications. It is hardly surprising, therefore, that they do not at the same time possess the most favourable properties for piezomagnetic applications.

Thus, for example, the commercial materials with the highest values of the saturation magnetostriction are very porous (see Table II), and consequently their maximum safe intensity and their piezomagnetic coupling are lower than would be the case for denser specimens. Fig. 11c shows the improvement in the coupling coefficient at remanence which is obtained by giving the materials the most favourable densities by special methods of preparation (e.g. higher sintering temperature). The coupling coefficient at remanence can be given still higher values by introducing slight changes in the chemical composition, namely by adding cobalt; we shall return to this point in the last section.

Table II. Saturation magnetostriction ( $\epsilon_s$ ), Curie temperature ( $T_c$  \*) and saturation polarization ( $J_s$ ) of nickel-zinc ferrites ( $\text{Ni}_{1-y}\text{Zn}_y\text{Fe}_2\text{O}_4$ ). Column 5 lists the corresponding commercial ferroxcubes, the approximate porosities of which are given under  $p$ ; the desired approximate porosities for maximum values of  $|k_{rem}|$  are given under  $p_1$ . When the material is completely free from pores its specific gravity is about 5.35 g/cm<sup>3</sup>.

$y$	$10^6 \epsilon_s$	$T_c$ (°C)	$J_s$ (Wb/m <sup>2</sup> )	Type of ferroxcube	$p$	$p_1$
0	-27	590	0.33 (1-p)	4E	0.25	0.12
0.20	-21	490	0.47 (1-p)	4D	0.20	0.09
0.35	-16	400	0.52 (1-p)	4C	0.15	0.07
0.50	-11	290	0.49 (1-p)	4B	0.11	0.06
0.65	-5	190	0.40 (1-p)	4A	0.08	0.05

\*) By Curie temperature is understood the temperature at which the saturation polarization becomes roughly zero, and not the temperature at which  $\mu$  sharply falls.

These improved materials will be put on the market as soon as demand for them warrants their production in reasonably large quantities.

In many cases, e.g. for high radiation intensities, it is not so much the highest possible value of the coupling coefficient at remanence that is required, as the highest possible value of  $|k|$  at the optimum biasing polarization,  $J_0 \approx 0.7 J_s$  (see fig. 4). The tensile strength should likewise be as large as

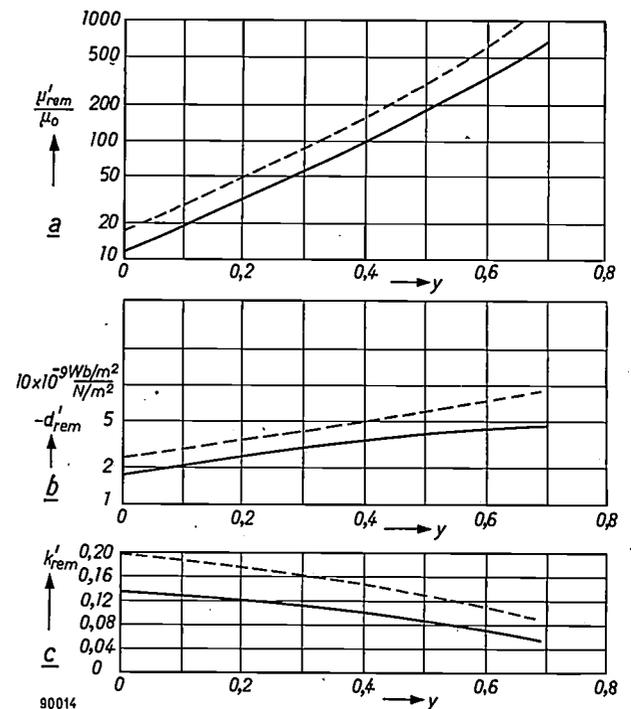


Fig. 11. Values at remanence of a) the relative permeability ( $\mu'_{rem}/\mu_0$ ), b) the sensitivity constant ( $d'_{rem}$ ), and c) the coupling coefficient ( $k'_{rem}$ ) for various types of ferroxcube 4 ( $\text{Ni}_{1-y}\text{Zn}_y\text{Fe}_2\text{O}_4$ ) as functions of the zinc-ferrite content  $y$ . Continuous line curves: commercial ferroxcube; dotted line curves: specially prepared types of ferroxcube of low porosity ( $p_1$  in Table II).

possible. Both requirements call for the highest attainable density of the ferrite; high density can be imparted either by carrying out the sintering process at a very high temperature or by replacing part of the nickel (and zinc) by copper <sup>17)</sup>.

The detection of acoustic waves

The half-wave window-type vibrators and ring vibrators shown in fig. 6, are also used for detecting acoustic waves. Another much-used form is the cylindrical receiver which may be regarded as a ring whose height is large compared with the diameter. This receiver is characterized by a sharp directivity pattern over a wide frequency range.

If it is required to receive a single frequency, the mechanical resonance frequency of the receiver must be made to coincide with that frequency. If a broad frequency band is to be detected, the resonance frequency is made higher than the highest frequency to be received, since then the frequency response is not so strongly dependent upon the frequency ( $\propto f$ ).

An important quantity in detection is  $(\partial B/\partial \sigma)_H = d$  introduced above, which determines the change in induction in the material brought about by an external stress.  $d$  is usually called the "sensitivity constant".

Nevertheless the quantity  $d$  does not always determine the sensitivity of the detector. This is connected with the fact that in the NiZn ferrite series (the ferroxcube 4 materials), the permeability  $\mu$  and likewise, if the number of turns  $n$  remains the same, the self-inductance  $L_w$  of the windings, increase with increasing  $d$  (fig. 11).

The determining quantity is the ratio of the amplitude  $\hat{V}$  of the open-circuit voltage to the amplitude  $\hat{\sigma}$  of the pressure of the incident wave:

$$\frac{\hat{V}}{\hat{\sigma}} = \frac{\omega n A \hat{B}}{\hat{\sigma}} = \omega n A |d|, \dots (9)$$

where  $A$  is the cross-sectional area.

The vibrational frequency of the circuit, the latter comprising the self-inductance  $L = L_w + L_s$  ( $L_s$  being the leakage inductance) and the capacitance  $C_0$  of the connecting cable etc.,

$$f_{cl} = \frac{1}{2\pi(LC_0)^{1/2}}, \dots (10)$$

may not fall below the value of the highest frequency for which (or up to which) the receiver is to be used. This is because at frequencies greater than  $f_{cl}$  the capacitance  $C_0$  would short-circuit the signal.

If  $L_s$  may be ignored,  $L_w$  ( $\approx L$ ) may be regarded as given. Since  $L_w$  is proportional to the product  $n^2\mu$ , the latter is fixed, so that  $n$  is proportional to  $\mu^{-1/2}$ . If we ignore the leakage, we find that, according to (9), the sensitivity of the detector is determined by  $d/\mu^{1/2}$ . This means that we should regard the coupling coefficient

$$k = d(E^H/\mu^\sigma)^{1/2}, \dots (11)$$

and not  $d$  as the determining quantity because  $E^H$ , the modulus of elasticity, is roughly the same for all dense NiZn-ferrites.

If  $L_s$  becomes a considerable fraction of  $L$ , as is the case in long, thin-walled cylindrical receivers (line type hydrophone), we are obliged to select a material for which  $\mu$  and likewise  $d$  are large (see fig. 11). In that case  $d$  is more or less the determining constant.

The relative variations of  $d$  and  $k$  as functions of  $B_0$  have been discussed earlier (see fig. 4). The absolute values of  $d$  and  $k$  for NiZn ferrites depend upon the zinc content (fig. 11). The latter shows that when  $k$  is the determining quantity, nickel ferrite is to be preferred to nickel-zinc ferrites.

For metals, e.g. "Permendur", the coupling coefficient at remanence  $|k_{rem}|$  has a value in the neighbourhood of 0.25. Such values can also be obtained with ferroxcube vibrators if the material is appropriately chosen (see the next section), so that the ferrites likewise bear comparison with metals for the detection of acoustic waves which is frequently done at remanence.

The influence of cobalt on the piezomagnetic properties of ferroxcube

In conclusion we shall discuss the application of ferroxcube filter elements, e.g. for band-pass filters. A band-pass filter may comprise a number of  $L$  and  $C$  elements and one or more vibrators. An example of such a filter is given in fig. 12. The aim of the filter is to pass with negligible attenuation a limited (usually not very narrow) frequency band while largely excluding frequencies outside that band. It is required, moreover, that the edges of the transmission band should be as far as possible unaffected by temperature variations.

The steepness of the edges is connected, amongst other things, with the width of the resonance peaks of the vibrators; that is to say with the mechanical quality factor  $Q = f/\Delta f$  ( $\Delta f$  is the interval

<sup>17)</sup> Y. Kikuchi and co-workers, Study on ferrites for the use in magnetostriction vibrators, Part II. Ni-Cu ferrite, Sci. Rep. Res. Insts. Tôhoku Univ. (B) 7, 171-178, 1955; see also: L. G. Van Uitert, Nickel-copper ferrites for microwave applications, J. appl. Phys. 27, 723-727, 1956.

between the two frequencies for which  $|Z|$  and  $|Y|$  (fig. 5) fall to very nearly  $1/\sqrt{2}$  of their maximum values). Laminated metals suffer from strong mechanical damping (frictional losses between the laminations) and therefore a correspondingly low

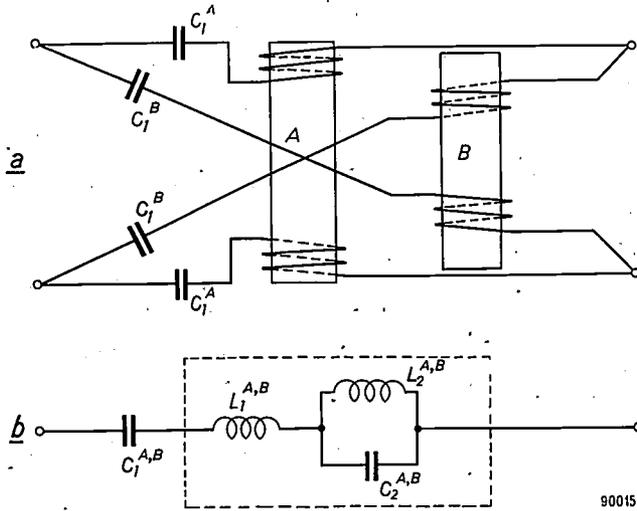


Fig. 12. a) Band-pass filter with ferrocube vibrators. Each of the two cores has two windings. Vibrators A and B differ slightly and have somewhat different resonance frequencies. (Taken from C.W. Diethelm's "Ferrite als magnetostruktive Resonatoren und deren Anwendung als Elemente elektrischer Filter", Dissertation E. T. H. Zürich 1951; see also Tech. Mitt. PTT 29, 281-297, 1951.) b) Equivalent circuit of one of the branches, magnetic and elastic losses being neglected.

$$L_1 \propto \mu^e, \quad L_2 = \frac{k^2}{1-k^2} L_1, \quad (L_2 C_2)^{-1/2} = 2\pi f^H.$$

quality factor  $Q$ . They are usually unsuitable for piezomagnetic filter elements. Ferrocube, on the other hand, which need not be laminated, possesses  $Q$  factors which are adequate for this application (between 2000 and 8000).

Temperature variation of the position of the edges is connected with the temperature variation of the anti-resonance frequencies  $f_a$  and the resonance frequencies  $f_r$  of the vibrators used in the filter and with the temperature variations of the permeability. For normal ferrites  $f_a$  and  $f_r$  fall off slowly as the temperature increases. The nickel used in the preparation of commercial nickel-zinc ferrites usually contains roughly half a percent of cobalt, so that the variation of  $f_a$  and  $f_r$  within the temperature range 20-50 °C is normally 0.1-0.25% (increasing with rising zinc content). For cobalt-free ferrites the variation is still greater (roughly 0.3%). All these variations are so large as to preclude the use of these ferrites in filters.

It has been found that the temperature dependence of  $f_a$ ,  $f_r$ , and the permeability of the various

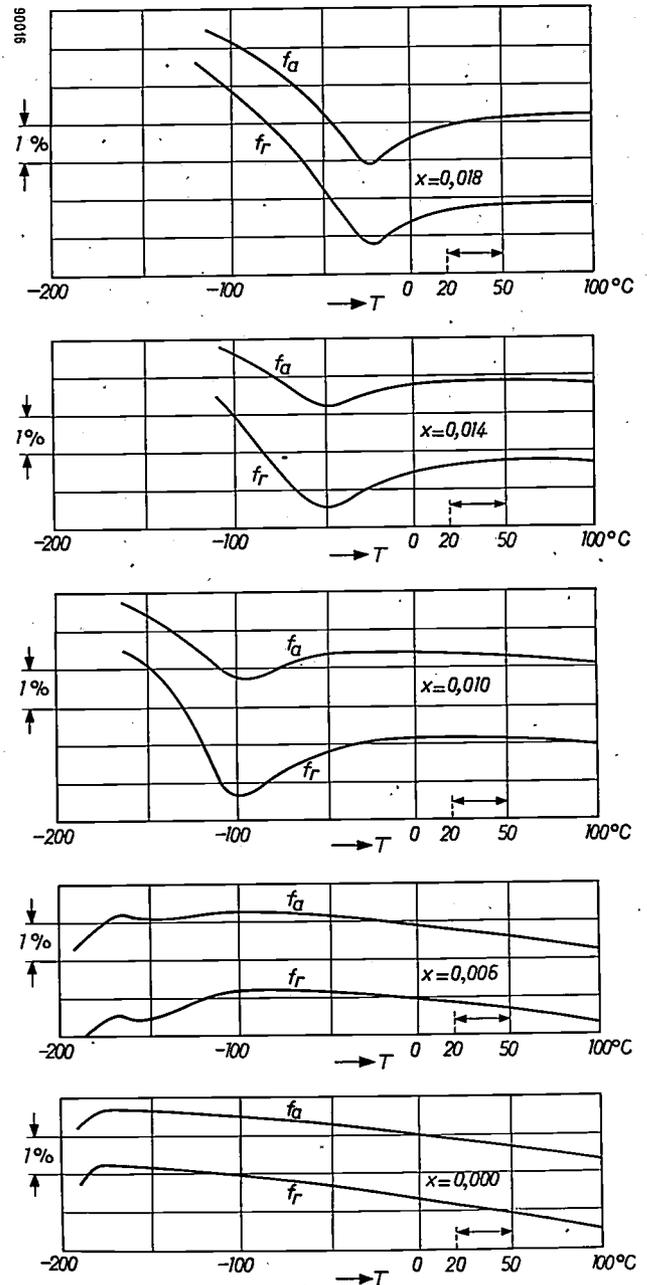


Fig. 13. Behaviour of  $f_a$  and  $f_r$  (see fig. 5) as functions of temperature for nickel ferrite to which various amounts of cobalt have been added ( $\text{Ni}_{1-x}\text{Co}_x\text{Fe}_2\text{O}_4$ ). Measurements at remanence.

ferrites may be considerably altered by adding more cobalt to the ferrite<sup>18) 19)</sup>.

In fig. 13 the percentage variations of  $f_a$  and  $f_r$  at remanence, for nickel-cobalt ferrites ( $\text{Ni}_{1-x}\text{Co}_x\text{Fe}_2\text{O}_4$ ), are plotted against the temperature; fig. 14 shows the initial permeability of the same ferrites

<sup>18)</sup> C. M. van der Burgt, Ferrites for magnetic and piezomagnetic filter elements with temperature-independent permeability and elasticity, shortly to appear in Proc. Inst. Electr. Engrs. (B) 104, 1957, Supplement No. 7.

<sup>19)</sup> C. M. van der Burgt, Controlled crystal anisotropy and controlled temperature dependence of the permeability and elasticity of various cobalt-substituted ferrites, shortly to appear in Philips Res. Rep. 12, 1957.

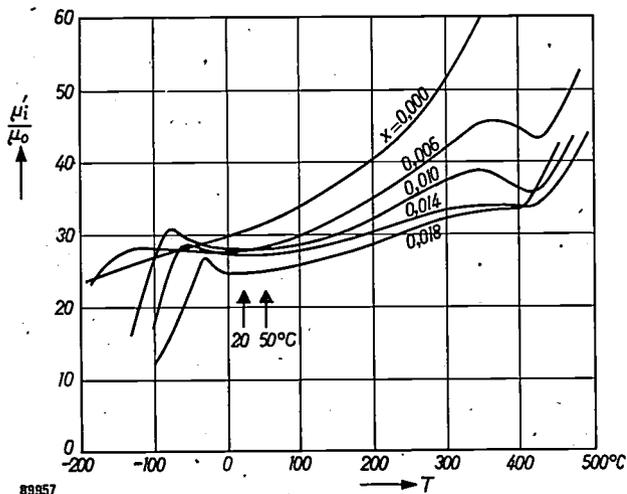


Fig. 14. Behaviour of the relative initial permeability ( $\mu_i'/\mu_0$ ) as a function of the temperature for nickel-cobalt ferrites ( $\text{Ni}_{1-x}\text{Co}_x\text{Fe}_2\text{O}_4$ ).

likewise as a function of temperature. It will be seen that a value for  $x$  can be found for which the variations of both  $f_a$  and  $f_r$  at remanence are less than 0.03% between 20 °C and 50 °C, while the variation of the reversible permeability at remanence is also strongly reduced. This is sufficient in many cases in practice. At a somewhat different value of  $x$  the variation of the initial permeability between 20 °C and 50 °C is a minimum.

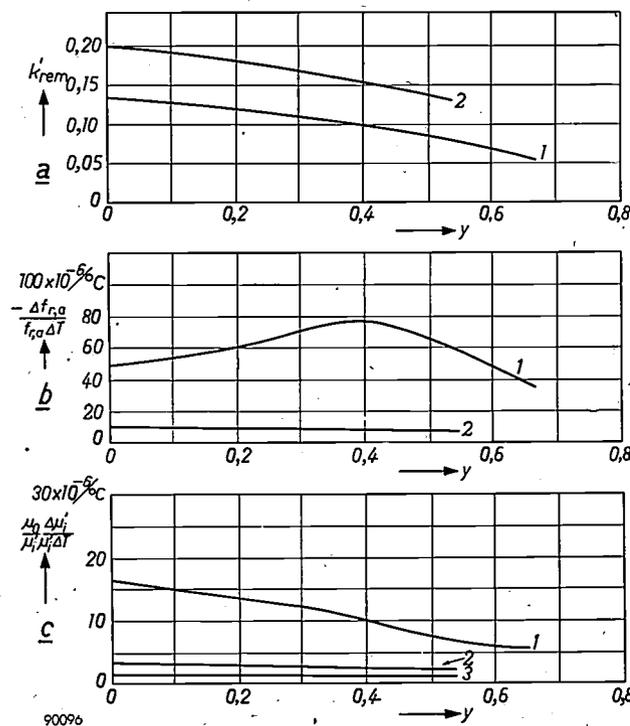


Fig. 15. Improvements in the temperature dependence (b and c) and the corresponding  $k'_{rem}$  (a) attained with  $(\text{Ni}_{1-y}\text{Zn}_y)_{1-x}\text{Co}_x\text{Fe}_2\text{O}_4$ ; curves 1 for commercial ferroxcube; 2 for ferroxcube with optimum cobalt content ( $x$ ) to obtain minimum values of  $-(\Delta f_{r,a}/f_{r,a}\Delta T)$  at remanence; 3 as 2 but now with  $x$  the optimum value for least temperature dependence of the relative initial permeability  $\mu_i'/\mu_0$ , i.e., for a minimum value of  $(\mu_0/\mu_i')(\Delta\mu_i'/\mu_i'\Delta T)$ .

A summary of results obtained with ring vibrators of  $(\text{Ni}_{1-y}\text{Zn}_y)_{1-x}\text{Co}_x\text{Fe}_2\text{O}_4$  (nickel-zinc-cobalt ferrite) is given in fig. 15.

The addition of cobalt likewise favours a high coupling coefficient. In fig. 16 the coupling coefficients for nickel-cobalt ferrites at remanence are plotted against the temperature. The figure shows that if the primary concern is for a large value of  $k$  at room temperature (under-water receivers and radiators), rather than for extreme temperature independence, nickel-cobalt ferrites can be prepared with a value of  $|k_{rem}|$  greater than 0.22. With somewhat higher densities a value of  $|k_{rem}|$  of about 0.25 can be attained. Given optimum initial polarization (see fig. 4),  $|k|$  can even be as high as 0.30.

Summarizing we can say that provided piezomagnetic ferroxcubes are specially prepared for the purpose, they can be successfully used instead of metals in ultrasonic vibrators. Another point in favour of these ferroxcube materials is that unlike the metals they involve only small amounts of

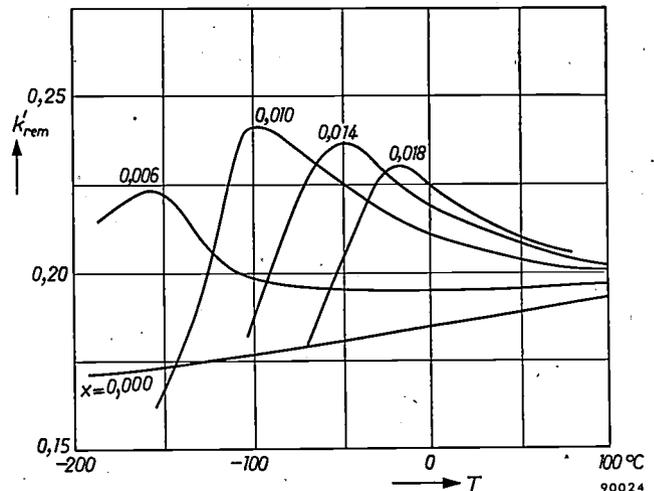


Fig. 16. Behaviour of  $k'_{rem}$  as a function of the temperature for nickel-cobalt ferrites ( $\text{Ni}_{1-x}\text{Co}_x\text{Fe}_2\text{O}_4$ ). The curves of this figure and those of figs 13 and 14 refer to rings with a porosity of 14 to 15%. For slightly smaller values of the porosity (10-12%; see also  $p_1$  in Table II), values of  $k'_{rem}$  of approximately 0.25 may be realized at room temperature.

expensive raw materials. In general the use of metals as piezomagnetic filter elements cannot be entertained because of their low mechanical  $Q$  (laminated vibrators) or because of their low piezomagnetic coupling (solid vibrators). Nickel-zinc ferrites can be used for the piezomagnetic filter elements provided the temperature dependence is sufficiently reduced by adding cobalt.

**Summary.** Piezomagnetic transducers have long been used for generating and detecting ultrasonic waves (e.g. in water). Various types of ferroxcube may be used as core material in place of piezomagnetic metals and alloys. Ferroxcube cores

need not be laminated. A short résumé is given of the piezomagnetic quantities that are relevant in this connection, and is followed by an account of the generation of piezomagnetic vibrations and a description of one or two conventional forms of transducer. The most important factors in the generation of acoustic waves are the electroacoustic efficiency and the maximum safe radiation intensity. For vibration frequencies above 20 kc/s the electroacoustic efficiency (measured at the optimum frequency) is higher for ferroxcube than for metals, since there are no eddy currents and since, in the absence of laminations, frictional losses are very low. The maximum safe intensity can be at least 10 W/cm<sup>2</sup> for continuous loading. In order to attain this high value, materials of good homogeneity and large saturation magnetostriction are required. Moreover, such materials require a high tensile strength; this may be achieved by suitable sintering techniques or suitable variations of the chemical compositions, which renders them less porous. High values for the piezomagnetic coupling coefficient, the

mechanical  $Q$  and the magnetic  $Q$  enhance the efficiency and therefore also the maximum safe intensity, since they make for smaller thermal stresses at a given power. For pulsed operation, which avoids thermal stresses, maximum safe intensities of as much as 15-20 W/cm<sup>2</sup> can be expected.

In general the suitability of ferroxcube as a material for acoustic detectors is also determined by the above-mentioned coupling coefficient.

Unlike metals, ferroxcube vibrators can be used in electric filters, when the latter call for both a high mechanical  $Q$  (2000 to 8000) and a high coupling coefficient. This application allows only a slight variation in the resonance and anti-resonance frequencies with temperature. By adding cobalt this variation can be kept within the desired limits (0.03%) for the required temperature range (20-50 °C). The addition of cobalt likewise improves the coupling coefficient, which in this way can be raised to a value at remanence nearly equal to that of the best metals in use at present (0.25).

## RECTANGULAR CATHODE-RAY TUBES OF HIGH ASPECT RATIO

621.385.832

For cathode-ray tube displays which are essentially long and narrow, e.g. special type radar ("A" scan) displays, there is considerable waste of equipment space when a circular-face tube is

used. Further, when several displays are observed by one operator, the distances between the individual displays are inconveniently large. A special tube with a long and narrow rectangular face (bulb

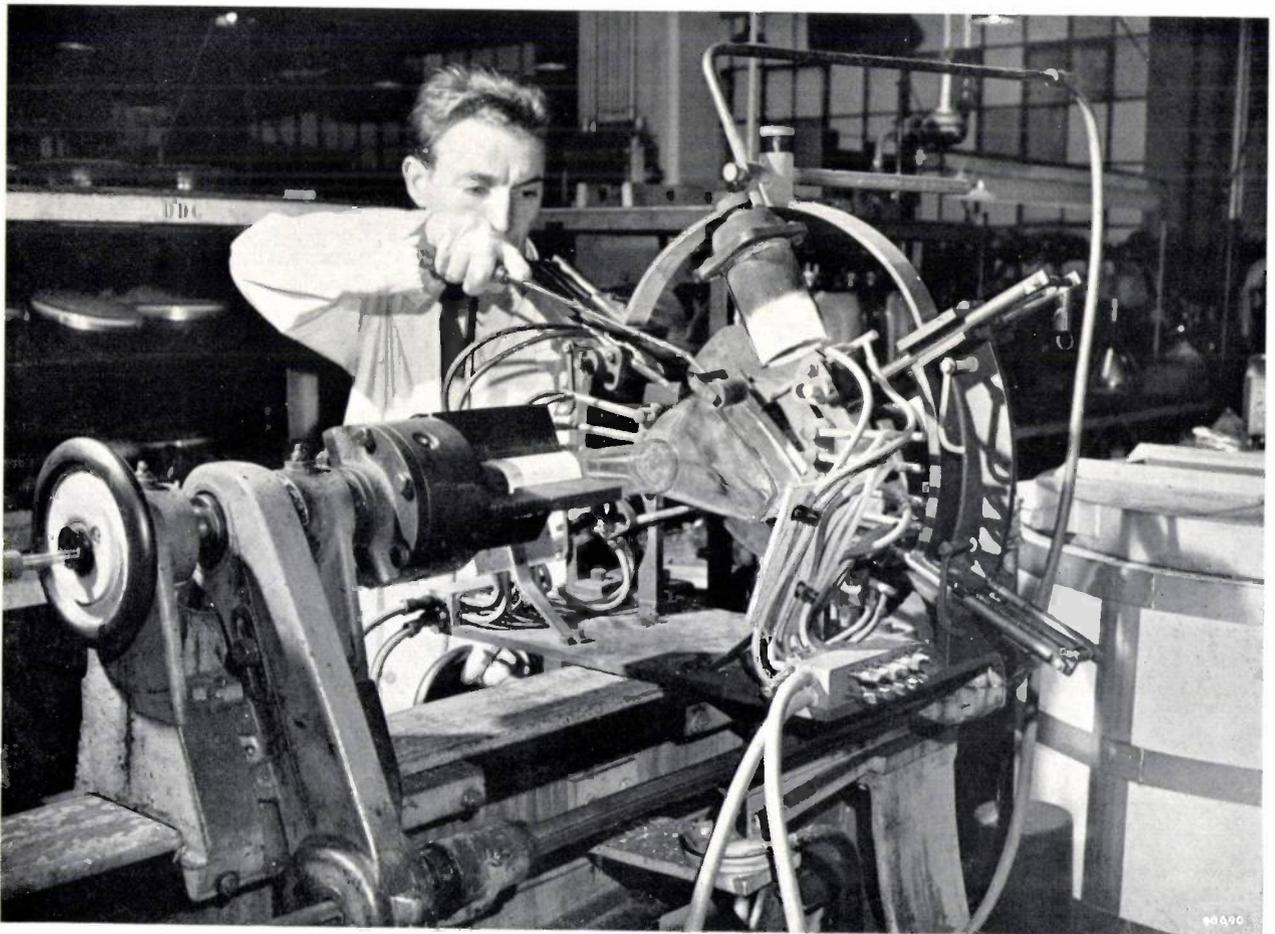


Fig. 1. Glass lathe for joining the pressed screen and cone of the rectangular cathode-ray tube bulb. The lathe contains eight stationary and four reciprocating burners. The latter are actuated by the large cam visible behind the bulb. The photograph shows the screen-cone joint during the rolling process.

need not be laminated. A short résumé is given of the piezomagnetic quantities that are relevant in this connection, and is followed by an account of the generation of piezomagnetic vibrations and a description of one or two conventional forms of transducer. The most important factors in the generation of acoustic waves are the electroacoustic efficiency and the maximum safe radiation intensity. For vibration frequencies above 20 kc/s the electroacoustic efficiency (measured at the optimum frequency) is higher for ferroxcube than for metals, since there are no eddy currents and since, in the absence of laminations, frictional losses are very low. The maximum safe intensity can be at least 10 W/cm<sup>2</sup> for continuous loading. In order to attain this high value, materials of good homogeneity and large saturation magnetostriction are required. Moreover, such materials require a high tensile strength; this may be achieved by suitable sintering techniques or suitable variations of the chemical compositions, which renders them less porous. High values for the piezomagnetic coupling coefficient, the

mechanical  $Q$  and the magnetic  $Q$  enhance the efficiency and therefore also the maximum safe intensity, since they make for smaller thermal stresses at a given power. For pulsed operation, which avoids thermal stresses, maximum safe intensities of as much as 15-20 W/cm<sup>2</sup> can be expected.

In general the suitability of ferroxcube as a material for acoustic detectors is also determined by the above-mentioned coupling coefficient.

Unlike metals, ferroxcube vibrators can be used in electric filters, when the latter call for both a high mechanical  $Q$  (2000 to 8000) and a high coupling coefficient. This application allows only a slight variation in the resonance and anti-resonance frequencies with temperature. By adding cobalt this variation can be kept within the desired limits (0.03%) for the required temperature range (20-50 °C). The addition of cobalt likewise improves the coupling coefficient, which in this way can be raised to a value at remanence nearly equal to that of the best metals in use at present (0.25).

## RECTANGULAR CATHODE-RAY TUBES OF HIGH ASPECT RATIO

621.385.832

For cathode-ray tube displays which are essentially long and narrow, e.g. special type radar ("A" scan) displays, there is considerable waste of equipment space when a circular-face tube is

used. Further, when several displays are observed by one operator, the distances between the individual displays are inconveniently large. A special tube with a long and narrow rectangular face (bulb

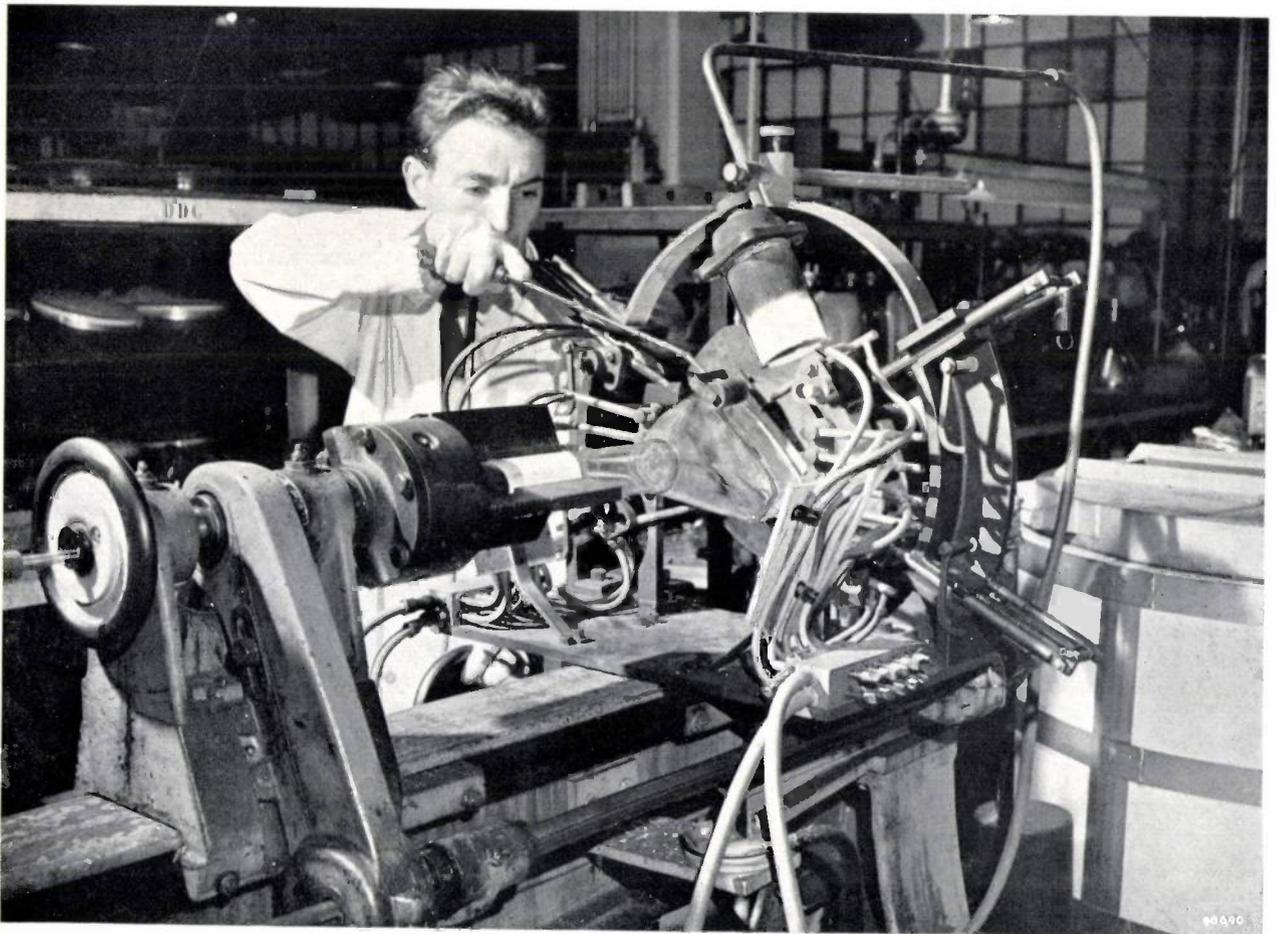


Fig. 1. Glass lathe for joining the pressed screen and cone of the rectangular cathode-ray tube bulb. The lathe contains eight stationary and four reciprocating burners. The latter are actuated by the large cam visible behind the bulb. The photograph shows the screen-cone joint during the rolling process.

periphery  $156 \times 66$  mm, useful screen area  $125 \times 35$  mm, i.e. an aspect ratio 3.5 : 1), has been developed for these cases. Tubes of this type permit of compact stacking and thus avoid both the disadvantages mentioned above. Some manufacturing problems not encountered in normal cathode ray tube production had to be solved for these tubes; a number of these problems will be briefly reviewed in this article.

Since it is impracticable to produce a strong enough bulb of the required shape by the usual process of blowing, the bulb design was based on the use of a pressed glass screen and cone. The depth of both screen and cone was chosen so as to permit satisfactory pressing notwithstanding their flat, narrow form; the sides of both are slightly curved for increased strength and slightly tapered for easy removal from the moulds. These requirements set a lower limit on the overall bulb dimensions for a given useful screen area.

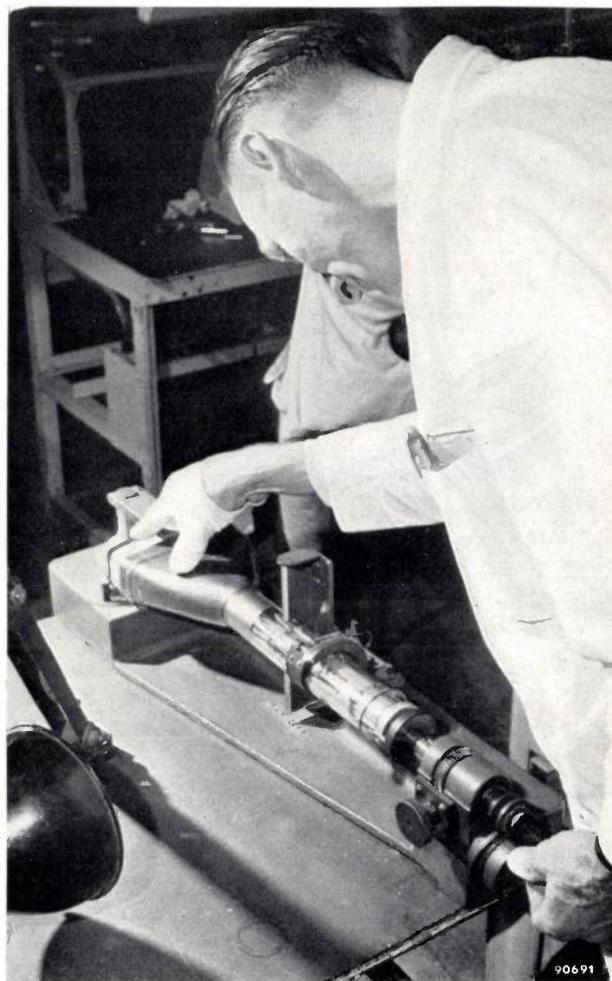
The joints between screen and cone and between cone and neck (a length of standard 50 mm tubing) are made on a glass lathe of special design. This lathe, shown in *fig. 1*, contains four reciprocating gas-oxygen burners in addition to the usual stationary ones. The stationary burners alone would heat the corners of the glass parts less than their sides, owing to the greater peripheral speed of the former. The reciprocating burners, which are actuated by a large cam on the head stock of the lathe (see figure), increase the heating on the corners only, and produce a uniform temperature round the edges of the glass.

The phosphor coating for the screen is produced by the well-known sedimentation technique<sup>1</sup>). For applying the internal graphite coating, a special brush was designed whose head can be swept across the flat internal surfaces of the bulb by manipulation at its long handle.

Electrostatic deflection is usually used in tubes for this type of display. The electron gun, including the deflector plates, has to be sealed into the neck with the proper orientation, i.e. with the direction of the scan (*x*-trace) accurately parallel to the long axis of the rectangular screen. Setting of the gun in the correct position is effected by means of a jig, making use of the fact that the *x*-trace deflector plates deflect the electron beam in a direction perpendicular to the axial slot of the "interplate shield" (the latter separates the two sets of deflector plates electrostatically in order to prevent "cross-talk"

between *x*- and *y*-deflection<sup>2</sup>)). The jig is shown in *fig. 2* and the aligning procedure is explained in the caption of the figure. By this method it has been possible to maintain a tolerance of  $\pm 2^\circ$  on the angle between the *x*-trace and the long axis of the screen.

Before sealing the glass foot carrying the gun into the neck of the cathode ray tube, the foot must be



*Fig. 2.* Jig for adjusting the electron gun in the tube neck so that the direction of the *x*-trace is parallel to the long axis of the rectangular screen. The screen end of the bulb is positioned unambiguously on a heavy casting whilst the neck rests in a semi-circular support: these two locations provide a reference plane defined by the axis of the tube and the long axis of the face. At the nominal position of the interplate shield of the electron gun assembly is a bracket painted with a white reference line (broken to allow space for the tube neck). The line has the same width as the interplate shield slot and is located in a plane perpendicular to the above-mentioned reference plane. The gun assembly, which fits tightly in the neck in order to prevent its vibrating out of position in subsequent handling, is moved up the neck by the mechanism seen in the lower right part of the photograph, until the interplate shield is at the position of the reference line. The gun is then rotated until the shield slot is seen in the same straight line as the reference line. The *x*-deflection will then be parallel to the former reference plane and hence to the long axis of the screen.

<sup>1</sup>) See the illustrations in Philips tech. Rev. 9, 339, 1947/48 and 10, 306, 1948/49.

<sup>2</sup>) J. de Gier and A. P. van Rooy, Improvements in the construction of cathode-ray tubes, Philips tech. Rev. 9, 180-184, 1947/48.

brought up to about 200 °C. This is usually done before insertion of the assembly into the neck, but since the aligning operation cannot be made with a hot foot, the pre-heating for the tubes described must take place with the gun and foot already in place. This is accomplished by a hot turret of steel placed around the stem. Nitrogen is passed up the stem of the foot into the bulb to prevent oxidation of the gun components. The nitrogen is pre-heated in order to avoid cracking of the hot foot.

Conventional methods are used in the further manufacturing processes such as pumping, ageing, etc.

The most important characteristics of the tube are the following: final anode voltage 5 kV, nominal screen current 10  $\mu$ A with a grid drive of 17.5 V, total light intensity 0.15 candela (owing to the use of a willemite fluorescent screen of very high efficiency); nominal deflection sensitivity in the *x*-direction 0.19 mm/V, in the *y*-direction 0.21 mm/V. The line width when scanning nearly the whole

screen area (describing an ellipse of 100  $\times$  30 mm) with optimum overall focus and nominal beam intensity is 0.5 mm. This good definition, which was part of the specification for the radar applications of the tube, was obtained by a special design of the electron gun: the undeflected spot size was reduced by a high cathode loading, a high first anode voltage (1800 V) and a long focussing anode (cf. the article quoted in <sup>2</sup>); the deflection astigmatism was diminished as far as possible by using relatively long deflection plates with a rather large separation. Striking a suitable compromise between length and separation of the plates was facilitated by the very limited scan in the *y*-direction. The relatively large deflection-plate assembly called for a neat mechanical construction in order not to increase the required tube neck diameter beyond the value of 50 mm mentioned above.

F. G. BLACKLER\*).

\* Mullard Radio Valve Co., Ltd., Mitcham (Surrey), England.

## FLOODLIGHTING OF THE DUTCH NATIONAL WAR MEMORIAL IN AMSTERDAM

628.974.6.064:725.94:621.327.534.2

A National War Memorial situated in the Dam Square in Amsterdam was unveiled by Queen Juliana of the Netherlands in May 1956.

The illumination of the 22 metres high column with its statuary and the low wall that serves as background is effected by means of 30 spot-lights disposed on the roofs of neighbouring buildings. This arrangement of the sources of light is in accord with the wishes of the sculptor John Rådecker and the architect Dr. J. J. P. Oud, that the character of the monument seen by day, be as far as possible preserved when the monument is illuminated by night.

To ensure reasonable intensity of illumination at the unavoidably large distances involved (40 to 100 metres), it was necessary to concentrate as much as possible of the light from the spot-lights on to the object. On the other hand, the high luminous intensity of the light beams could not be allowed to interfere with traffic in the busy square nor could light be allowed to shine into the windows of the buildings opposite, so that the light beams had to be strictly confined to the object to be floodlit. In view of the slender form of the object it proved that the problem could be solved simply and economically by using high-pressure mercury-vapour lamps.

The lamps used are of the HPK 125 W type.

The discharge is concentrated in the axis of the tube, and is about 30  $\times$  1.5 mm in size; the mean brightness of this very elongated source is 640 stilb. The vertically disposed source is projected on the column of the monument, with a magnification of several hundred times, by means of a spherical mirror. The mirror has a sealing cover-plate which has horizontal flutes to elongate the image of the light source in a vertical direction, the source being less slender in shape than the monument. Weaker light from the outer edge of the discharge and from the quartz wall, which would pass wide of the monument, is intercepted by screens of ceramic material which are fitted close up against the discharge tube to avoid penumbras. The HPK-type lamps were chosen with an eye to this requirement: in this type of lamp the discharge tube is exposed to the air<sup>1</sup>) and not mounted in an external envelope as in normal HP or HPL lamps. To ensure that the screens remain precisely in position when the tube is renewed, the construction shown in the diagram (p. 302) is employed.

The luminous flux produced by the 30 spot-lights with their total power of about 4 kW is approximately 140 000 lumens; the luminance of the monument is 10 cd/m<sup>2</sup>.

<sup>1</sup>) See: W. Elenbaas, Philips tech. Rev. 18, 167-172, 1956/57.

brought up to about 200 °C. This is usually done before insertion of the assembly into the neck, but since the aligning operation cannot be made with a hot foot, the pre-heating for the tubes described must take place with the gun and foot already in place. This is accomplished by a hot turret of steel placed around the stem. Nitrogen is passed up the stem of the foot into the bulb to prevent oxidation of the gun components. The nitrogen is pre-heated in order to avoid cracking of the hot foot.

Conventional methods are used in the further manufacturing processes such as pumping, ageing, etc.

The most important characteristics of the tube are the following: final anode voltage 5 kV, nominal screen current 10  $\mu$ A with a grid drive of 17.5 V, total light intensity 0.15 candela (owing to the use of a willemite fluorescent screen of very high efficiency); nominal deflection sensitivity in the *x*-direction 0.19 mm/V, in the *y*-direction 0.21 mm/V. The line width when scanning nearly the whole

screen area (describing an ellipse of 100  $\times$  30 mm) with optimum overall focus and nominal beam intensity is 0.5 mm. This good definition, which was part of the specification for the radar applications of the tube, was obtained by a special design of the electron gun: the undeflected spot size was reduced by a high cathode loading, a high first anode voltage (1800 V) and a long focussing anode (cf. the article quoted in <sup>2</sup>); the deflection astigmatism was diminished as far as possible by using relatively long deflection plates with a rather large separation. Striking a suitable compromise between length and separation of the plates was facilitated by the very limited scan in the *y*-direction. The relatively large deflection-plate assembly called for a neat mechanical construction in order not to increase the required tube neck diameter beyond the value of 50 mm mentioned above.

F. G. BLACKLER\*).

\* Mullard Radio Valve Co., Ltd., Mitcham (Surrey), England.

## FLOODLIGHTING OF THE DUTCH NATIONAL WAR MEMORIAL IN AMSTERDAM

628.974.6.064:725.94:621.327.534.2

A National War Memorial situated in the Dam Square in Amsterdam was unveiled by Queen Juliana of the Netherlands in May 1956.

The illumination of the 22 metres high column with its statuary and the low wall that serves as background is effected by means of 30 spot-lights disposed on the roofs of neighbouring buildings. This arrangement of the sources of light is in accord with the wishes of the sculptor John Rådecker and the architect Dr. J. J. P. Oud, that the character of the monument seen by day, be as far as possible preserved when the monument is illuminated by night.

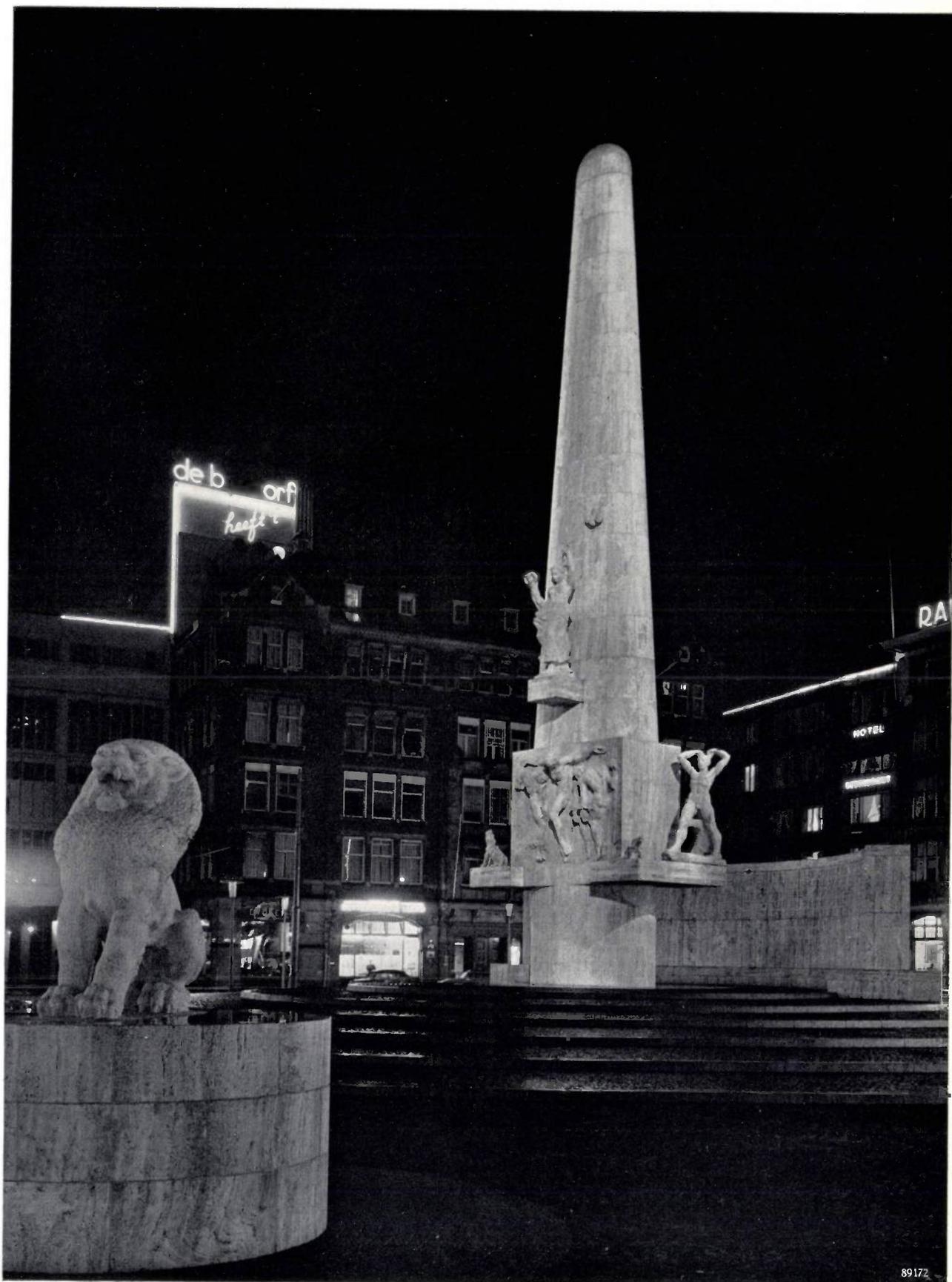
To ensure reasonable intensity of illumination at the unavoidably large distances involved (40 to 100 metres), it was necessary to concentrate as much as possible of the light from the spot-lights on to the object. On the other hand, the high luminous intensity of the light beams could not be allowed to interfere with traffic in the busy square nor could light be allowed to shine into the windows of the buildings opposite, so that the light beams had to be strictly confined to the object to be floodlit. In view of the slender form of the object it proved that the problem could be solved simply and economically by using high-pressure mercury-vapour lamps.

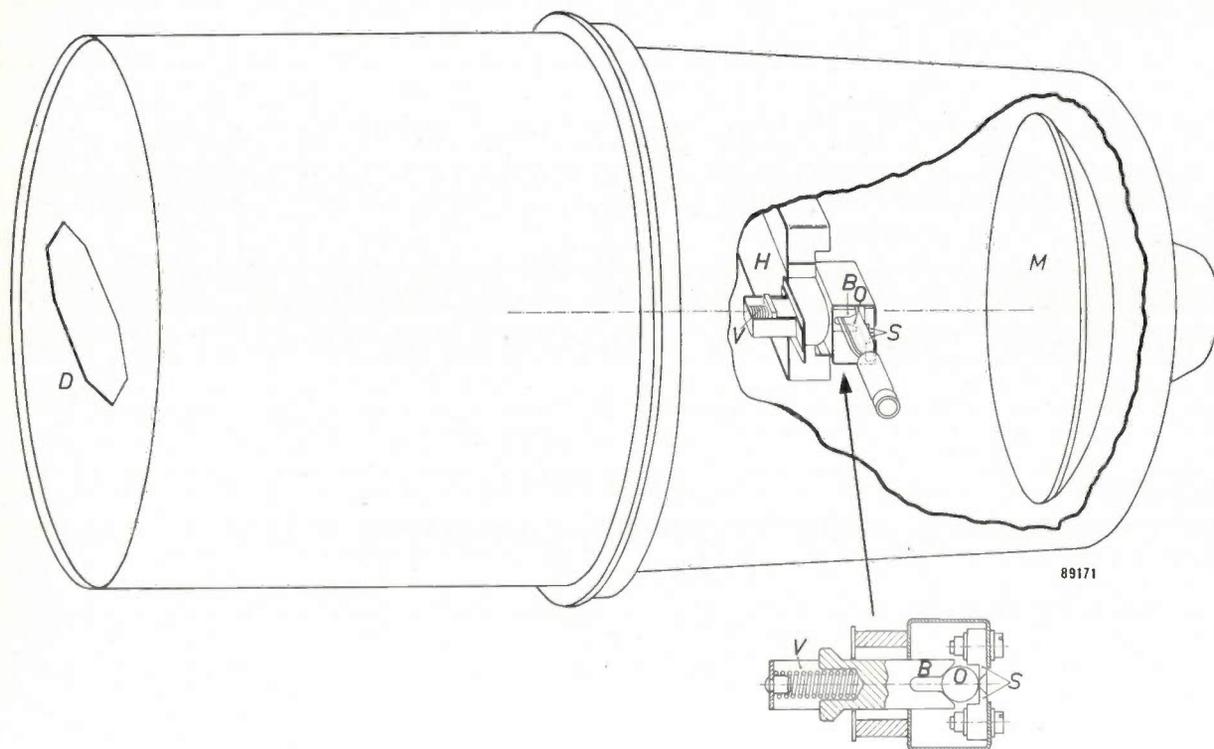
The lamps used are of the HPK 125 W type.

The discharge is concentrated in the axis of the tube, and is about 30  $\times$  1.5 mm in size; the mean brightness of this very elongated source is 640 stilb. The vertically disposed source is projected on the column of the monument, with a magnification of several hundred times, by means of a spherical mirror. The mirror has a sealing cover-plate which has horizontal flutes to elongate the image of the light source in a vertical direction, the source being less slender in shape than the monument. Weaker light from the outer edge of the discharge and from the quartz wall, which would pass wide of the monument, is intercepted by screens of ceramic material which are fitted close up against the discharge tube to avoid penumbras. The HPK-type lamps were chosen with an eye to this requirement: in this type of lamp the discharge tube is exposed to the air<sup>1</sup>) and not mounted in an external envelope as in normal HP or HPL lamps. To ensure that the screens remain precisely in position when the tube is renewed, the construction shown in the diagram (p. 302) is employed.

The luminous flux produced by the 30 spot-lights with their total power of about 4 kW is approximately 140 000 lumens; the luminance of the monument is 10 cd/m<sup>2</sup>.

<sup>1</sup>) See: W. Elenbaas, Philips tech. Rev. 18, 167-172, 1956/57.





Spotlight constructed for the floodlighting of the National Monument. It is shown here turned somewhat about its axis (the discharge tube and the aperture are not vertical) and with part of the outer case cut away to reveal the HPK lamp. The discharge tube *O* of the HPK lamp rests in a ceramic block *B* which is pressed against the ceramic screens *S* by a spring *V*. This "lamp holder" is attached to a bracket *H*, which is mounted in the spotlight and can be accurately adjusted once and for all with respect to the mirror *M* of the optical system by means of screws. The cylindrical case of this spotlight with its front aperture *D* intercepts any light that might be emitted outside the desired beam.

In order to show the floodlighting to its best possible advantage, the public lighting in the square around the monument has been altered, so that there are no bright sources of light to distract the passer-by. This has been done without conferring a lifeless character on the façades of the surrounding buildings.

The 1/50 scale model of the square (shown below) in which the monument is situated ("the Dam") was

employed to good purpose in preliminary experiments relating to the modification of the street lighting and the floodlighting of the monument. These experiments were undertaken in the laboratory of the Municipal Gas and Electricity Works, Amsterdam.

M. RUTGERS van der LOEFF \*) and J. B. de BOER.

\*) Chief Engineer of the Municipal Gas and Electricity Works, Amsterdam.



## SEALING AMPOULES OF LYOPHYLIZED VACCINE



The photograph shows the sealing of ampoules containing lyophilized vaccine, i.e. vaccine which has been frozen in ampoules and dried in vacuo in the frozen state. Such vaccines are used both in medical and in veterinary practice.

## A LOUDSPEAKER INSTALLATION FOR HIGH-FIDELITY REPRODUCTION IN THE HOME

by G. J. BLEEKSMAN and J. J. SCHURINK.

62L.395.623.8

---

*For more than a quarter of a century the normal broadcast receiver has been equipped with a single loudspeaker fitted inside the cabinet. The introduction of sets with two separate speakers, each reproducing part of the frequency spectrum on the Philips "Bi-Ampli" principle, dates only from the last few years.*

*The installation described in this article goes a big step further, the loudspeakers being entirely separated from the amplifying part; two low-note speakers are together housed in a special cabinet and two high-note speakers separately in their own boxes. The quality of reproduction has been remarkably improved in this way, which appears to full advantage with F.M. reception and the play-back of gramophone records or tape recordings.*

---

A complete system for sound reproduction, from performance to play-back can be considered as a chain whose first link is formed by one or more microphones and whose final link is made up of one or more loudspeakers. In between are one or more amplifiers, and further either a radio link or some "memory" with play-back facilities. (The "memory" is normally a tape recording or a gramophone record.)

The links in this chain are not equally strong. "Strong" links are those that are unlikely to be hardly improved upon by present engineering practice. As such we may consider the condenser microphone and the amplifiers. The memory is among the weaker links. However high the quality of magnetic recordings<sup>1)</sup> and modern gramophone records<sup>2)</sup> may be, certain improvements in the step function response and in the dynamic characteristic are still desirable.

Loudspeakers too should be considered as among the weaker links. A recent development in this field forms the subject of this article. The installation described is in our opinion about the very best attainable for home entertainment that can be realized with the technical means now at our disposal.

### Brief description of the installation

The installation comprises four loudspeakers, viz. two for the low notes (with a range below 420 c/s) and two for the frequency range above 420 c/s (referred to here briefly as the high range). The

two low-note speakers are together housed in a corner cabinet (*fig. 1*), whilst the high-note speakers are contained in separate boxes (*fig. 2*). All these loudspeakers are fed by one amplifier (type AG 9000 or AG 9006), via a low-pass and a high-pass filter respectively.

The loudspeakers for the frequency range above 420 c/s radiate nearly all their energy in a forward direction, and for this reason they are called projectors. Separate projectors can be so placed that their sound reaches the listeners only indirectly, i.e. via one or more reflections from the walls or the ceiling of the room. The importance of this method is that it diffuses the sound, so that the result approaches far more closely the actual conditions in the concert hall. Measurements in various concert halls have revealed that at most places in the hall the greater part of the sound intensity is attributable to indirect sound and only a small part to direct radiation<sup>3)</sup>. Similar conditions can be created in the living-room by directing the projectors at different walls of the room in such a way that the listeners are outside the direct beams (*fig. 3*). The position of the low-note cabinet is less critical, since the directions from which the low notes issue are more difficult to distinguish.

The use of two projectors, placed some distance apart from each other and from the listeners, effectively avoids the "hole-in-the-wall" effect<sup>4)</sup>. This can be explained as follows. Let us assume that we are listening to an orchestra via one loudspeaker only. Even if the whole reproduction channel reproduced uniformly the entire frequency spectrum and were completely devoid of distortion and

<sup>1)</sup> See e.g. D. A. Snel, Philips tech. Rev. 14, 181-190, 1952/53, and W. K. Westmijze, Philips tech. Rev. 15, 84-96, 1953/54.

<sup>2)</sup> See L. Alons, Philips tech. Rev. 13, 134-144, 1951/52, and J. L. Ooms, Philips tech. Rev. 17, 101-109, 1955/56.

<sup>3)</sup> Cf. Philips tech. Rev. 17, 258, 1955/56.

<sup>4)</sup> Philips tech. Rev. 17, 171, 1955/56.

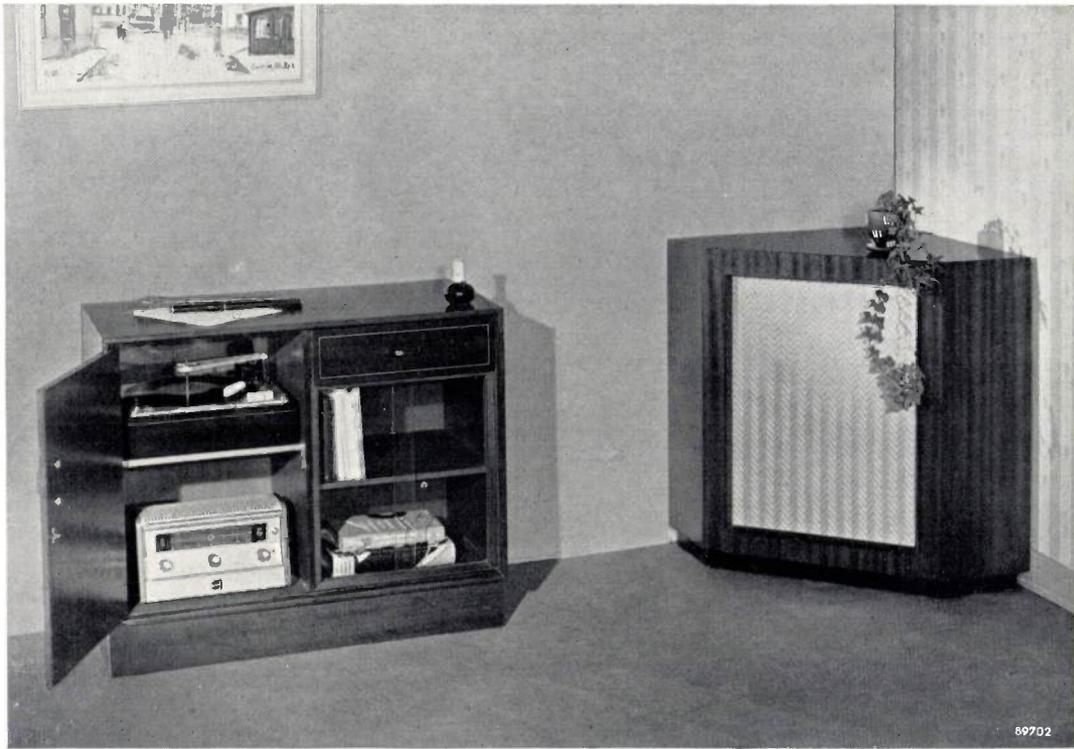


Fig. 1. Left: gramophone AG 1005 (with magnetodynamic pick-up AG 3020/21<sup>5)</sup>) and amplifier AG 9006. Right: the low-note cabinet. The installation is completed by two separate loudspeakers for the high notes, above 420 c/s (see figs. 2 and 3).

dynamic limitations, the reproduction would still remain unsatisfactory in one respect: the entire orchestra, set up on a wide podium, would be heard as if it were compressed into the small aperture of the loudspeaker, i.e. as if the orchestra was listened to through a hole in the wall of the concert hall.

An effective means of remedying this hole-in-the-

wall effect is stereophony<sup>6)</sup>, but this is only seldom used in radio broadcasts or with gramophone records. However, by placing the two projectors in such a way that the sound reaches the listeners only indirectly, the hole-in-the-wall effect can be eliminated.

The high-note speakers (type 9710 M) are distinguished from the low-note speakers (type 9710) in that they are double-cone loudspeakers<sup>7)</sup>. As can be seen in *fig. 4*, the moving coil drives not only a normal cone ( $C_1$ ) but also a small auxiliary cone ( $C_2$ ). The latter extends the frequency range, which would normally reach no further than 8 kc/s, to about 18 kc/s, and at the same time blurs the sharp outlines of the beam.

The reproduction of the high notes is further improved by the use of loudspeakers with nearly constant impedance: at 20 kc/s the impedance of a type 9710 M loudspeaker is only 1.5 times as high as at 400 c/s, whereas for conventional loudspeakers this ratio is at least 5. When fed with a constant voltage, the current flowing through such "constant-

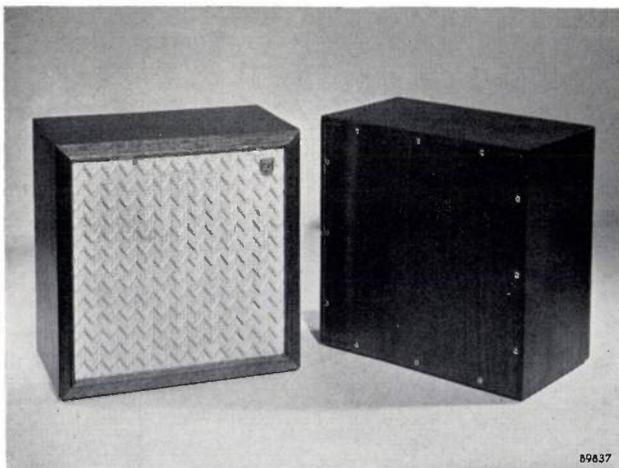


Fig. 2. The two high-note loudspeakers. Each box contains one double-cone loudspeaker and has a width and height of 26 cm.

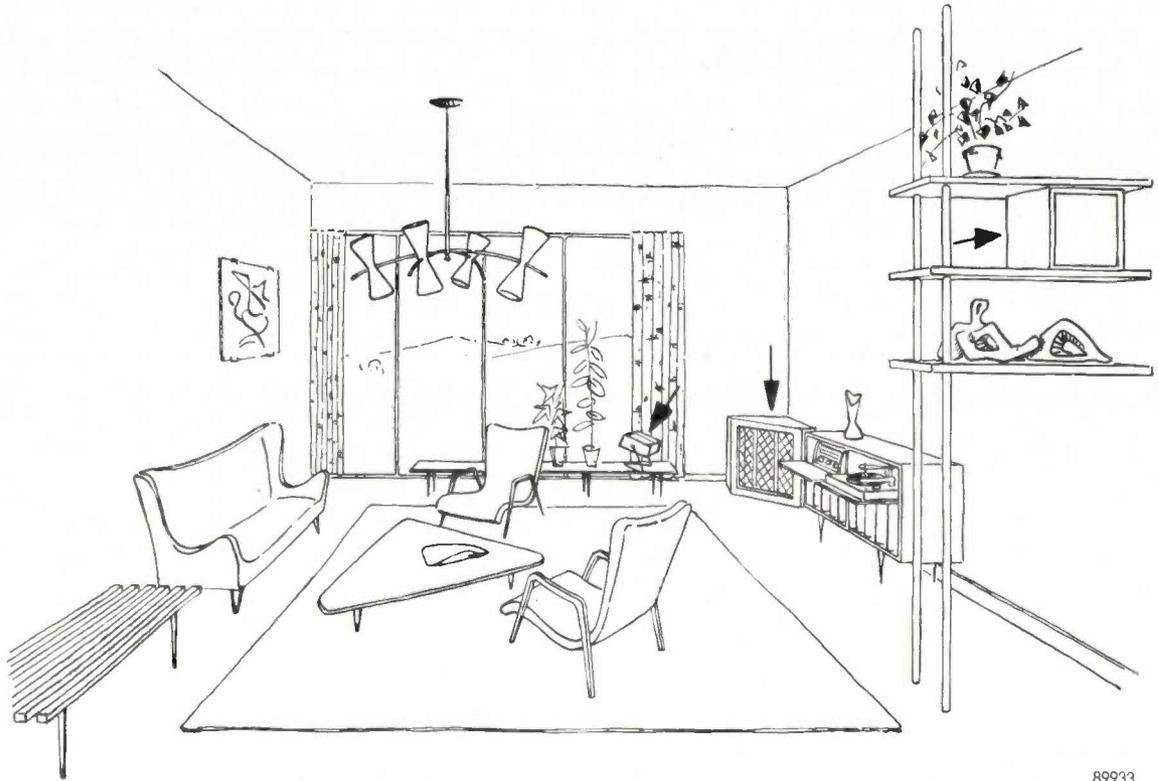
<sup>5)</sup> N. Wittenberg, Philips tech. Rev. 18, 101-109 and 173-178, 1956/57.

<sup>6)</sup> See e.g. the article quoted in <sup>4)</sup>, p. 173 et seq.

<sup>7)</sup> J. J. Schurink, Philips tech. Rev. 16, 241-249, 1954/55.

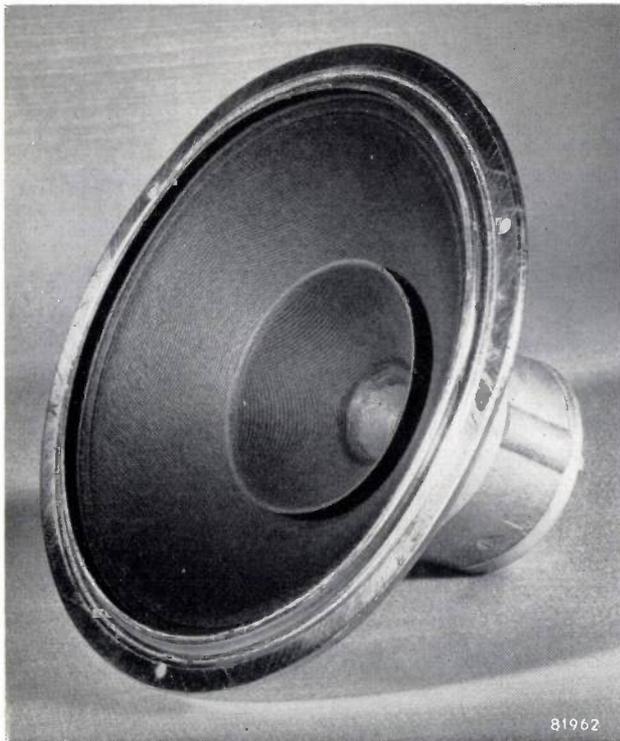
impedance" loudspeakers will therefore decrease considerably less with rising frequency than the current flowing through conventional loudspeakers.

The reproduction of the high-note range will consequently be superior, since it is the current that determines the force with which the cone is driven.



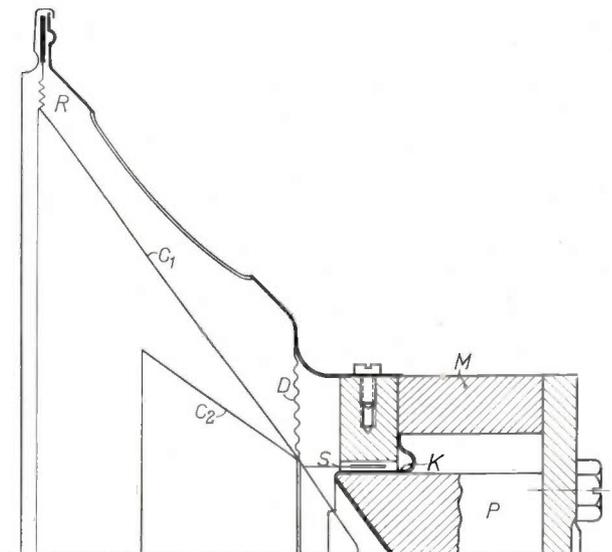
89933

Fig. 3. A typical arrangement of the units. To the right is the cabinet containing the gramophone and the amplifier, and in the corner the low-note cabinet. The two high-note projectors are so arranged that a diffused sound is obtained.



81962

Fig. 4. The double-cone loudspeaker 9710 M of the high-note projector. In the cross-section,  $C_1$  is the main cone,  $C_2$  the auxiliary inner cone,  $S$  the coil,  $D$  the centring ring,  $R$  the corrugated cone edge,  $M$  the permanent magnet, and  $K$  the copper ring ensuring a constant impedance (see further in this article). The core  $P$  of the magnetic circuit has a conical hole for accommodating the top of cone  $C_1$ .



The cabinets in which the loudspeakers are housed are completely closed behind the cone. As regards the high-note projectors this prevents any back radiation (possibly directed at the listeners) which would mar the diffusion effect aimed at. As regards the low-note loudspeakers the closed box prevents the air vibrations from travelling round the cone; it thus represents an infinitely large baffle, which greatly enhances the reproduction of the very low notes. The enclosed volume of air, as we shall demonstrate later in this article, also helps in reducing non-linear distortion. To attain the best possible enclosure, "complete cones" have been used instead of the conventional truncated cone (fig. 4).

After this brief description, we shall now consider some of the above-mentioned features more closely, starting with the forms of distortion caused by non-linear phenomena.

**Non-linear distortion in loudspeakers**

The speech coil of a moving-coil loudspeaker, through which a current  $I$  flows, is situated in a magnetic field with an induction  $B$ . A force  $F = BIl$  then operates on the coil ( $l$  denoting the length of wire in the speech coil). Since a force is numerically equal to the reaction it produces, this force is equal to the product of the mechanical impedance  $Z_m$  and the velocity  $v$ :

$$BIl = Z_m v = \left( j\omega M + \frac{1}{j\omega C} + R_m \right) v. \quad (1)$$

In this expression  $\omega$  is the angular frequency of the current,  $M$  the mass of the coil and cone and that of the air that moves with the cone,  $C$  the total compliance (i.e. the reciprocal of the stiffness) and  $R_m$  the mechanical resistance (both the resistance of the effective sound radiation and the mechanical loss resistance, the latter mainly occurring in the edge of the cone and in the centring ring).

The displacement  $x$  of the coil at the moment  $t$  is given by  $x = \int v dt$ , and hence

$$x = \frac{v}{j\omega} = \frac{BIl}{\frac{1}{C} - \omega^2 M + j\omega R_m} \dots (2)$$

The resonance frequency  $f_0$  of the system is therefore:

$$f_0 = \frac{1}{2\pi\sqrt{MC}} \dots (3)$$

Because of the resistance term in (2), the displacement will be greatest at a frequency below  $f_0$ .

It also appears from (2) that at frequencies that are sufficiently above the resonance frequency the displacement varies about proportionally with  $\omega^{-2}$ , and hence by 12 dB per octave (the current remaining constant). In this frequency range the amplitude is so small that no distortion is noticeable. At frequencies sufficiently below the resonant frequency, however, the displacement is given by

$$x \approx CBIl \dots (4)$$

Here the amplitude can become so large that the relation between  $x$  and  $I$  is no longer a linear one as the result of two causes: 1) the induction  $B$  in the air-gap is not quite homogeneous, so that  $B$  depends upon  $x$  (at very large amplitudes the coil comes even partly out of the air-gap), and 2) the compliance  $C$  is dependent on  $x$  to an even higher degree. With a sinusoidal current this non-linear distortion will produce harmonics, so that the timbre of the sound will be different. Moreover, if the current consists of two (or more) sinusoidal components, non-harmonic overtones are likely to occur, which are displeasing to the ear. We shall now separately consider these two forms of distortion.

*Harmonic distortion*

It can be seen from fig. 5a that the factors  $B$  and  $C$  in (4) are not constants. The dot-dash lines of this diagram represent the force  $F$  (plotted horizontally) operating on the coil in the magnetic field, at different values of a direct current  $I$  passing through the coil. The force  $F$  is strictly proportional to  $I$ , but varies with  $x$ , since  $B$  varies with the position in the air-gap. If the field were homogeneous, all these lines would be straight and vertical. If the displacement  $x$  of the currentless coil is measured statically as a function of an external force  $F$ , we arrive at a hysteresis loop as represented by the fully drawn curve in fig. 5a. This shows that the total compliance  $C_c$  of the cone edge and the centring ring (the only compliances operative here) depends upon  $x$  and thus likewise has a non-linear character.

The variation of  $x$  as a function of time, for a sinusoidal current, can be derived from the curves in fig. 5a. As expected, the original sinusoidal waveform is considerably distorted and a number of higher harmonics occur in the displacement. As regards the sound pressure  $p$ , this effect will be even more pronounced since the pressure variations are proportional to  $\omega^2 x$ , so that the subsequent harmonics in  $p$  are respectively 1, 4, 9, ... times greater than those in  $x$ . For instance, 1% of third harmonic in  $x$  results in 9% of third harmonic in  $p$ .

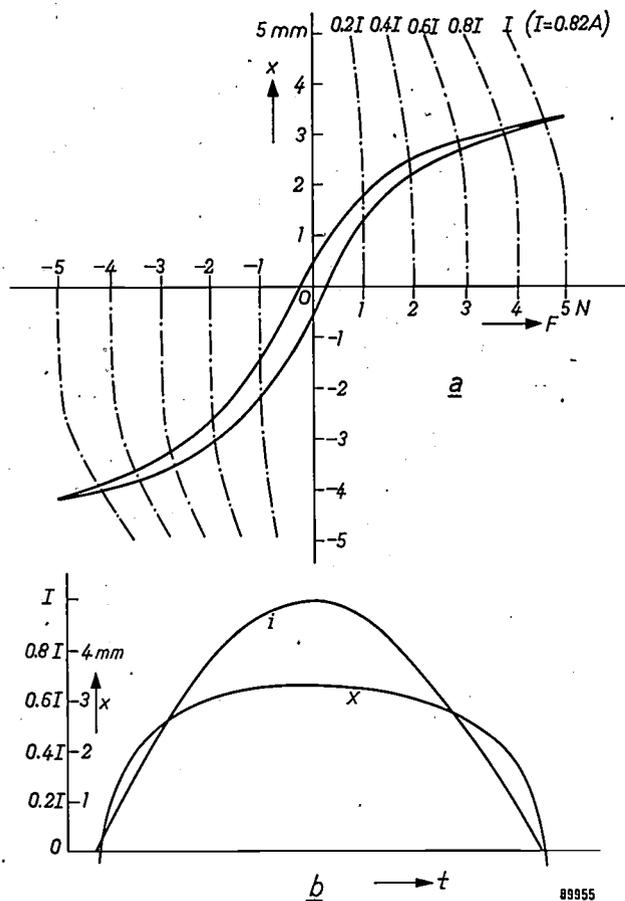


Fig. 5. a) Dot-dashed curves: force  $F$  on the coil of a moving-coil loudspeaker with a direct current of  $1/5I$ ,  $2/5I$ , ...  $I$  passing through the coil ( $I = 0.82 A$ ). Full curve: static displacement  $x$  of the (currentless) coil as a function of an externally applied force  $F$ .  
 b) Displacement  $x$  as a function of time  $t$ , showing distortion as compared with the sinusoidal current  $i$  passed through the speech coil.

**Non-harmonic distortion**

Let us now consider the case that the current is made up of two sinusoidal components, so that its instantaneous value  $i$  can be written as:

$$i = \hat{I}_1 \sin 2\pi f_1 t + \hat{I}_2 \sin 2\pi f_2 t \dots \quad (5)$$

The non-linear relation between displacement  $x$  and current  $i$  can be expressed by a series:

$$x = x_0 + x_1 i + x_2 i^2 + x_3 i^3 + \dots \quad (6)$$

Substituting (5) in (6) shows that the displacement will contain components with the frequencies  $\pm m f_1 \pm n f_2$  ( $m$  and  $n$  being integers 0, 1, 2, ...). This effect is known as intermodulation.

The above is only valid for frequencies below the resonance frequency, in the so-called stiffness range, in which the amplitude is independent of the frequency. Above the stiffness range the amplitude does depend upon the frequency, which greatly complicates a quantitative evaluation. Qualitatively speaking, however, intermodulation appears here as well.

Non-harmonic distortion can be attributed to still another cause, viz. the Doppler effect. If a loudspeaker cone vibrates simultaneously with a low frequency  $f_1$  and a higher frequency  $f_2$ , then it may be considered as a source of sound with frequency  $f_2$  whose distance to the ear varies with the frequency  $f_1$ . Owing to the Doppler effect the pitch will then fluctuate with the rhythm of the low frequency  $f_1$  around the higher frequency  $f_2$ , so that frequency modulation occurs. This can be interpreted as the presence of a sound with central frequency  $f_2$  and two side-bands in which the frequencies  $f_2 \pm m f_1$  occur. Let the amplitudes of the vibrations with the frequencies  $f_1$  and  $f_2$  be  $\hat{x}_1$  and  $\hat{x}_2$  respectively, then the amplitudes of the first and the second side-band components will amount to

$$10^{-2} \times \hat{x}_1 f_2 \% \text{ and } 5 \times 10^{-7} \times (\hat{x}_1 f_2)^2 \% \text{ of } \hat{x}_2,$$

$\hat{x}_1$  being expressed in cm and  $f_2$  in c/s. For example, if  $\hat{x}_1 = 0.1$  cm and  $f_2 = 10\,000$  c/s we find: 10% for the components with the frequencies  $f_2 + f_1$  and  $f_2 - f_1$ , and 0.5% for the components with the frequencies  $f_2 + 2f_1$  and  $f_2 - 2f_1$ .

**Means of reducing distortion**

*Doubling the number of loudspeakers*

For reducing the non-linear distortion there exists a solution so simple that its value is not always fully appreciated. We refer to the use of more than one loudspeaker (quite apart from the separation of low and high notes). To attain an equivalent total sound volume each individual cone can then vibrate with a smaller amplitude, so that a shorter and less curved part of the hysteresis loop (fig. 5a) is used, and the coil operates within a more uniform magnetic field. For this reason we are using two low-note loudspeakers in our installation (the two high-note speakers are used mainly to diffuse the sound, distortion at higher frequencies being anyway very slight).

As for the low-note reproduction, this particular aim — less distortion through smaller amplitude, at equal total sound volume — might also be attained in principle by using instead of two adjacent loudspeakers with radius  $a$  (fig. 6c), a single loudspeaker with radius  $a\sqrt{2}$  (fig. 6b).

Let us first compare one "small" loudspeaker (radius  $a$ , fig. 6a) with one "large" loudspeaker (radius  $a\sqrt{2}$ , fig. 6b). Their radiation resistances are proportional to the fourth power of their radii. The large speaker, therefore, has a radiation resistance  $(\sqrt{2})^4 = 4$  times greater than that of the small loudspeaker. At the same amplitude  $\hat{x}$  the large loudspeaker will radiate four times more power.

When the large loudspeaker is to be compared with a combination of two small ones then one would expect, if each small speaker, operating separately, radiates a power  $P$ , that they would produce together a power  $2P$ , and that therefore the large loudspeaker, with  $4P$ , would be twice as powerful. It

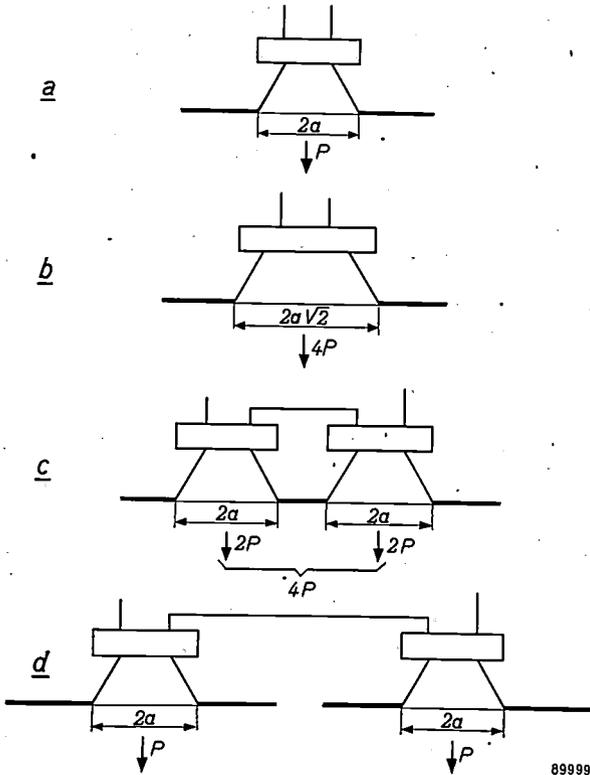
should be taken into account, however, that the effect of two small loudspeakers close together is different from that of the same two speakers some distance apart. Two loudspeakers (vibrating in phase) in close proximity produce the favourable effect that they stimulate each other, as it were, to radiate a greater power.

This effect can be readily explained as follows; if the loudspeakers are far apart (fig. 6d), then each yields energy to the air, since the cone is moving against self-produced pressure variations. If, however, the loudspeakers are close together (fig. 6c), then in addition to the self-produced pressure variations there will be those from the adjacent loudspeaker. Owing to the small distance, these pressure variations are virtually equal in amplitude and phase. Each of the cones, then, operates against a sound pressure that is twice as high and thus produces double the amount of power. The two loudspeakers together now produce a power that is twice as large as when they are placed far apart (for a constant amplitude  $\hat{x}$  of the cone displacement).

Klapman has evaluated this effect for intermediate cases, viz. for two loudspeakers, considered as flat pistons (radius  $a$ , distance between centres  $d$ )<sup>8</sup>. Fig. 7a shows the curve plotted for  $R_r/\rho c A$  ( $R_r$  = radiation resistance per piston,  $\rho$  = air density,  $c$  = velocity of sound,  $A$  = area of each piston) as a function of  $2\pi fa/c$ , with  $d$  as parameter. For the type 9710 loudspeaker  $2a = 17.5$  cm; the frequency scales added to fig. 7 apply to this value. In the cabinet the loudspeakers are placed at a distance  $d = 2.8 a$ . In the frequency range in question (below 420 c/s)  $R_r$  is substantially greater than would be the case with a single loudspeaker (curve for  $d = 0$ ).

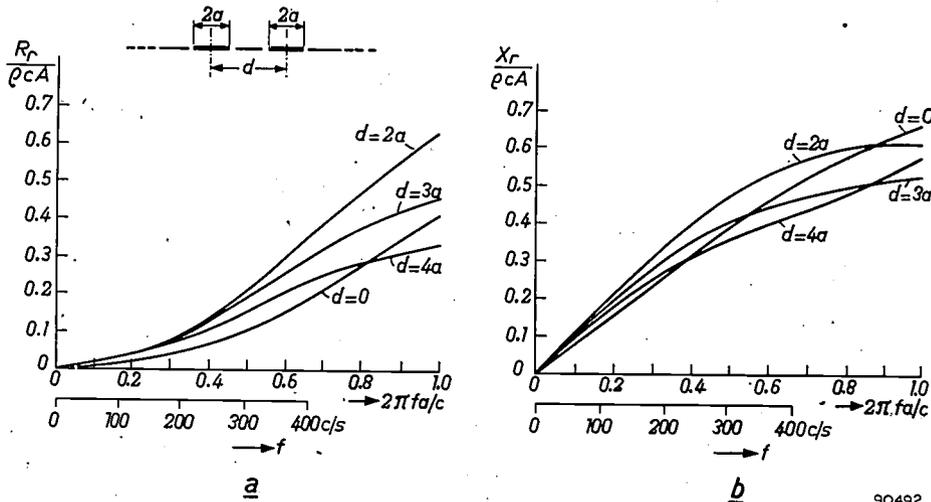
In the above imaginary experiment, in which two loudspeakers are brought closer together, the amplitude  $\hat{x}$  was assumed to remain constant. If, however, the supply current is kept constant when the loudspeakers are approaching each other,  $\hat{x}$  will decrease a little (thus reducing the advantage), because an increasing amount of air will move with the cones. This moving mass of air gives rise to a reactive component ( $X_r$ ) of the mechanical impedance. The curve of this component has likewise been evaluated by Klapman (fig. 7b).

<sup>8</sup> S. J. Klapman, Interaction impedance of a system of circular pistons, J. Acoust. Soc. Amer. **11**, 289-295, 1939/40.



89999

Fig. 6. a) Loudspeaker with diameter  $2a$ . The radiated sound power is  $P$ .  
 b) Loudspeaker with diameter  $2a\sqrt{2}$ . Sound power  $4P$ .  
 c) Two loudspeakers as in (a), placed closely together. Now either radiating a sound power  $2P$ , they produce together  $4P$ .  
 d) Two loudspeakers as in (a), placed a distance apart. As in (a), either produces a power  $P$ ; total power  $2P$ .  
 (The loudspeakers are of course supposed to be of the same type, mounted in large baffles, and vibrating with the same amplitude  $\hat{x}$ .)



90492

Fig. 7. a) The quantity  $R_r/\rho c A$ , and b) the quantity  $X_r/\rho c A$ , plotted against  $2\pi fa/c$ , for two circular pistons in an infinite baffle, vibrating in phase (according to Klapman<sup>8</sup>). The frequency scale is that for a diameter  $2a = 17.5$  cm (loudspeakers 9710).

A large cone, however, has the drawback that it must be relatively heavy to be sufficiently rigid. The loudspeakers used here (type 9710), with  $2a = 17.5$  cm, are preferable in this respect to a loudspeaker with  $2a = 17.5\sqrt{2} = 25$  cm, the cone of which is more than twice as heavy.

*Loudspeakers in closed cabinets*

A second means of reducing distortion is to enclose a given volume of air behind the cone. The enclosed air represents a stiffness  $S_a$ , which, unlike the rigidity  $S_c$  of cone edge and centring ring, remains satisfactorily constant and thus tends to improve the linearity of the entire system.

In order to find to what extent  $S_a$  still deviates from true linearity, we shall consider an enclosed quantity of air under a piston. If the latter is given a displacement  $\Delta x$ , then the air rigidity  $S_a$  exerted on the piston can be written as:

$$S_a \approx \frac{\rho_0 c^2 A^2}{V_0} \left[ 1 + \frac{1}{2}(\kappa + 1) \frac{A}{V_0} \Delta x \right] \dots (7)$$

(the process being assumed as adiabatic). Here  $\rho_0$  represents the density and  $V_0$  the volume of the air in the state of equilibrium,  $A$  the area of the piston and  $\kappa$  the ratio  $c_p/c_v$  of the specific heat of air at constant pressure to that of air at constant volume ( $\kappa \approx 1.4$ ).

Formula (7) can be derived as follows. An adiabatic process occurs according to

$$pV^\kappa = \text{constant}$$

( $p$  = air pressure,  $V$  = air volume). For a displacement of the piston causing the pressure to increase from  $p_0$  to  $p_0 + \Delta p$  and the volume to decrease from  $V_0$  to  $V_0 - \Delta V$ , we arrive at

$$(p_0 + \Delta p) (V_0 - \Delta V)^\kappa = p_0 V_0^\kappa.$$

From this we obtain

$$\frac{\Delta p}{\Delta V} = \kappa p_0 \left[ V_0^{-1} + \frac{\kappa+1}{2!} V_0^{-2} \Delta V + \frac{(\kappa+1)(\kappa+2)}{3!} V_0^{-3} (\Delta V)^2 + \dots \right].$$

The stiffness  $S_a$  can be defined as  $\Delta F/\Delta x$ ,  $\Delta F$  being the force necessary to bring about a displacement  $\Delta x$  of the piston. Here  $\Delta F = A\Delta p$  and  $\Delta x = \Delta V/A$ , so that:

$$S_a = \frac{\Delta F}{\Delta x} = A^2 \frac{\Delta p}{\Delta V} = \frac{\kappa p_0 A^2}{V_0} \left[ 1 + \frac{\kappa+1}{2} \frac{A}{V_0} \Delta x + \frac{(\kappa+1)(\kappa+2)}{6} \frac{A^2}{V_0^2} (\Delta x)^2 + (\dots) (\Delta x)^3 + \dots \right]. (8)$$

With the help of the relation  $c = \sqrt{\kappa p_0/\rho_0}$ , the product  $\kappa p_0$  in the right-hand term of (8) can be written as  $\rho_0 c^2$ . By substituting this value and disregarding the terms containing  $(\Delta x)^2$ ,  $(\Delta x)^3$ , etc. in (8), we arrive at equation (7).

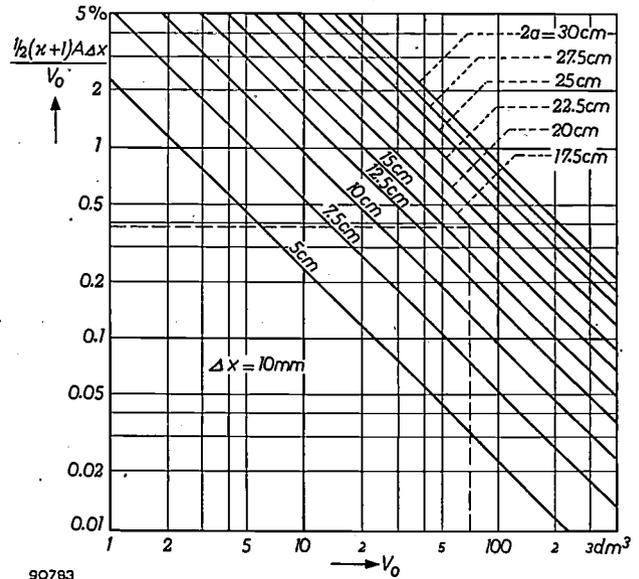


Fig. 8. The first non-linear term in the rigidity of an enclosed volume of air,  $\frac{1}{2}(\kappa + 1)A\Delta x/V_0$ , as a percentage of the main term, plotted against  $V_0$ , for various values of the cone diameter  $2a$ , for  $\Delta x = 10$  mm (see eq. 7).

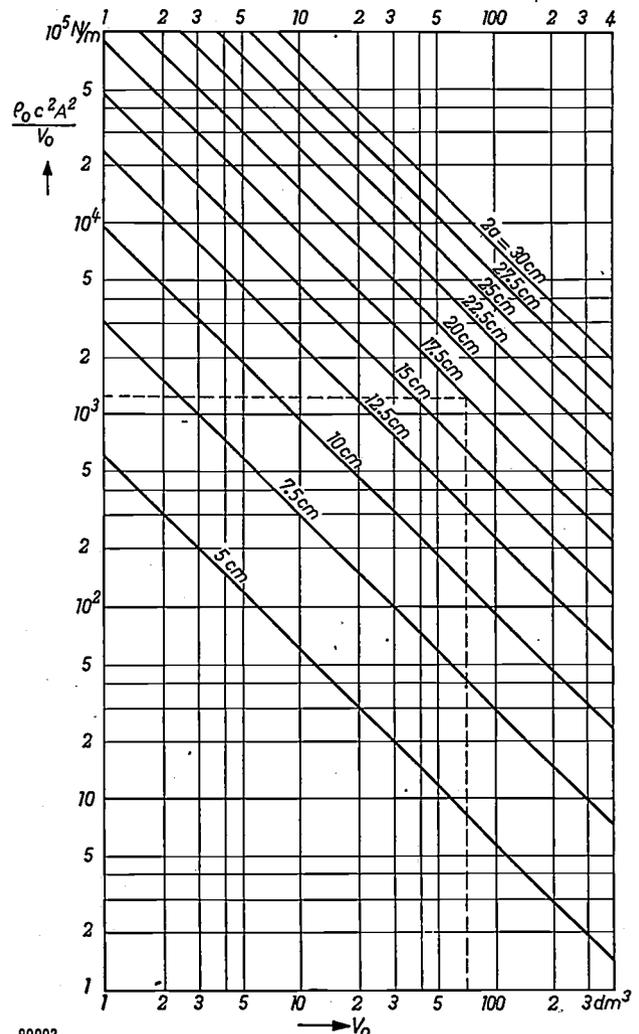


Fig. 9. Main term  $\rho_0 c^2 A^2/V_0$  of the rigidity of an enclosed volume of air (see eq. 7), versus  $V_0$ , for various values of the cone diameter  $2a$ .

The term  $\frac{1}{2}(\kappa + 1)A\Delta x/V_0$  represents the non-linearity of  $S_a$ . In *fig. 8* this term is expressed as a percentage and plotted as a function of the enclosed air volume  $V_0$ , with the diameter  $2a$  of the cone as parameter; this diagram applies to the extremely large amplitude  $\hat{x} (= \Delta x) = 10$  mm. It then appears that the non-linearity in  $S_a$  can be kept below 0.5% by a proper selection of  $2a$  and  $V_0$  (e.g.  $2a = 17.5$  cm, such as that of the loudspeaker 9710, and  $V_0 = 70$  l, which is half the volume of the low-note cabinet; this gives 70 l per loudspeaker).

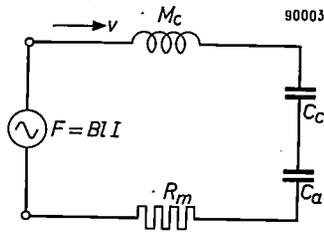


Fig. 10. Equivalent diagram of a loudspeaker with a cone compliance  $C_c$ , for a compliance  $C_a$  of the enclosed air volume behind the cone. The self-inductance  $M_c$  represents the effective mass of the cone and of the air moving with it; the resistance  $R_m$  the sum of the radiation resistance and the loss resistance.

Owing to this enclosure of a quantity of air behind the cone, the total stiffness  $S$  becomes  $S_c + S_a$ ; the non-linearity of the cone stiffness  $S_c$  appears from the hysteresis loop in *fig. 5a*. It is evident that because  $S_a$  is virtually linear,  $S$  will show relatively less distortion than  $S_c$ , and the less so as  $S_a$  is greater and therefore as the air volume  $V_0$  is smaller (see eq. (7) and also *fig. 9*, which represents the linear part of  $S_a$  as a function of  $V_0$  with  $2a$  as parameter).

The air rigidity cannot be indiscriminately enlarged since it is associated with an increase in the resonance frequency of the cone. This is revealed by (3), which now assumes the following form:

$$f_0 = \frac{1}{2\pi \sqrt{M \frac{C_c C_a}{C_c + C_a}}} = \frac{1}{2\pi \sqrt{\frac{M}{S_c + S_a}}}$$

( $C_c = 1/S_c$  and  $C_a = 1/S_a$  are the compliance values; cf. the equivalent diagram *fig. 10*). The resonance frequency lies only slightly below  $f_0$ , and the former must be confined to the lowest regions of the frequency range to be reproduced. The resonance frequency of the type 9710 loudspeakers without cabinet is about 40 c/s. The cabinet for the low-note speakers (*fig. 1*) is made so large (70 l per loudspeaker) that  $S_a$  considerably exceeds  $S_c$  ( $S_a = 1200$  N/m,  $S_c = 750$  N/m). This has raised the resonance frequency to about 60 c/s. By using electric damping (cf. the section *The amplifier* at the end of this article) this resonance peak has been sufficiently flattened out.

The high-note speakers are likewise accommodated in closed boxes; but here the object is to prevent back radiation. The boxes can be small (*fig. 2*) — and therefore the air rigidity great — since the resonance frequency of these speakers can quite permissibly be raised to about 300 c/s. In fact, a high resonance frequency is an advantage since it helps to suppress low notes, and so reinforces the effect of the electric filter, to be discussed presently.

*Division of the frequency range*

In the foregoing, intermodulation and Doppler effect were mentioned as causes of non-harmonic distortion. Both can be effectively combated by splitting up the audio range and reproducing it by separate loudspeakers.

It was found that in practice a division into two parts, as applied in the installation discussed here, is sufficient. As to the question at what frequency this division can best be made, it was decided to make the cross-over frequency 420 c/s. Combination tones of the type  $\pm mf_1 \pm nf_2$ , where  $f_1$  is low (e.g. 50 c/s), are especially objectionable if  $f_2$  is higher than about 400 c/s, so that the cross-over frequency should preferably not be fixed much higher than this value.

The filter effecting the separation of the high and low notes consists of two coils with self-inductance  $L$  and two capacitors of capacitance  $C$ , connected as shown in *fig. 11*. The impedance of each of the

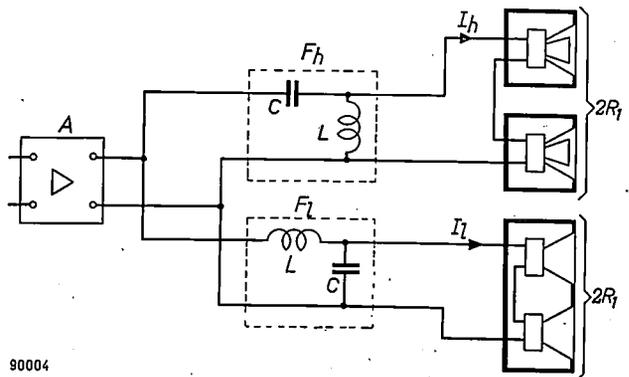


Fig. 11.  $F_l$  low-pass filter ( $f < 420$  c/s),  $F_h$  high-pass filter ( $f > 420$  c/s),  $A$  amplifier. For loudspeakers with a resistance  $R_1 = 7 \Omega$ ,  $L = 7.4$  mH and  $C = 20 \mu F$ .

speakers is a nearly frequency-independent resistance  $R_1$ , amounting to  $7 \Omega$  for the low-impedance speakers, and to 400 or 800  $\Omega$  for the high-impedance ones. (How this constant impedance has been achieved will be discussed later in this article.) Let the internal resistance of the amplifier be negligible, and  $E$  be its output voltage; we then find for the

current  $I_l$  through the low-note speakers and  $I_h$  through the high-note projectors:

$$I_l = \frac{E}{2(1 - \omega^2 LC)R_1 + j\omega L}$$

and

$$I_h = \frac{E}{2\left(1 - \frac{1}{\omega^2 LC}\right)R_1 - \frac{j}{\omega C}}$$

The formulae may be written as:

$$I_l = \frac{E/2R_1}{\sqrt{\left\{1 - \left(\frac{\omega}{\omega_0}\right)^2\right\}^2 + \left(\frac{\omega}{\omega_0}\delta\right)^2}} \quad (9)$$

and

$$I_h = \frac{E/2R_1}{\sqrt{\left\{1 - \left(\frac{\omega_0}{\omega}\right)^2\right\}^2 + \left(\frac{\omega_0}{\omega}\delta\right)^2}} \quad (10)$$

where

$$\omega_0^2 = \frac{1}{LC}$$

and

$$\delta = \frac{1}{2R_1} \sqrt{\frac{L}{C}} \quad (11)$$

$\delta$  being the damping factor.

In order to give each network the desired pass band (fig. 12),  $L$  and  $C$  must be so chosen that

$$\frac{1}{2\pi\sqrt{LC}} = 420 \text{ c/s}, \quad (12)$$

and there must also be adequate damping  $\delta$ . For critical damping (aperiodic system without free vibrations),  $\delta$  must be at least 2. In that case, however, the currents  $I_l$  and  $I_h$  are insufficiently

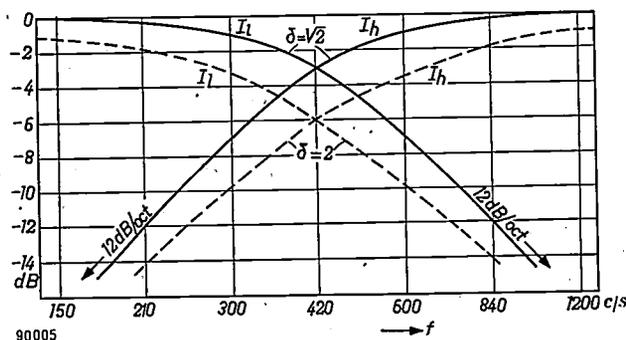


Fig. 12. Currents  $I_l$  and  $I_h$  through the low-note and high-note loudspeakers respectively, plotted against the frequency  $f$ , for a constant signal voltage and a negligible internal resistance of the amplifier, both scales being logarithmic. The dotted curves apply for critical damping ( $\delta_{cr} = 2$ ), the full curves for  $\delta = 1/\sqrt{2}$ .

constant in their pass bands; at the cross-over frequency they will drop by as much as 6 dB below the value  $E/2R_1$  which they respectively assume at very low and very high frequencies (see the dotted curves in fig. 12). For obtaining a flatter characteristic we have preferred a somewhat smaller damping, viz.  $\delta = 1/\sqrt{2}$ . The fully drawn curves in fig. 12, which drop only 3 dB below  $E/2R_1$  at the cross-over frequency, represent this damping value. Outside the pass bands the drop in the curves approaches 12 dB per octave, i.e.  $I_l$  decreases at high frequencies proportionally to  $f^{-2}$  and  $I_h$  increases at low frequencies proportionally to  $f^2$ , which is also apparent from the formulae (9) and (10).

If  $R_1$  is known,  $L$  and  $C$  are completely determined by (11), (12) and the value selected for  $\delta$ . For example, when  $R_1 = 7 \Omega$  we find:  $L = 7.4$  mH and  $C = 20 \mu\text{F}$ , which are the values used for low-impedance loudspeakers.

### Some further details regarding the cabinets

One advantage of operating the loudspeakers in completely enclosed cabinets is, as we demonstrated earlier, a reduction of distortion, as a result of the additional air rigidity. Another advantage, at least as regards the low notes, is the fact that the sound waves cannot travel around the cone (the same as if the baffle were infinitely large). This latter point may be elucidated as follows.

The sound radiation of a vibrating diaphragm is impaired if the air vibrations are allowed to travel along a short path around the diaphragm. This is why this path is usually lengthened by placing the diaphragm in a baffle<sup>9)</sup>, e.g. inside a cabinet. Let  $l_0$  be the shortest path from the front to the rear of the diaphragm around the baffle, then we may apply the rule of thumb that the sound emission decreases by 6 dB per octave if the frequency drops below the value relating to a wavelength  $\lambda = 2l_0$ . For a frequency of 50 c/s, for instance, the wavelength is more than 6 m. An enormous baffle would thus be required for a satisfactory reproduction of very low notes.

In the installation described here this has been circumvented by fitting the bass-note speakers in a completely enclosed cabinet which, just as an infinite baffle, prevents even the longest waves from traveling around.

With any cabinet, however, whether closed or not, there is a risk that it will act as a resonant cavity and give rise to standing waves due to reflection against

<sup>9)</sup> See e.g. Th. van Urk and R. Vermeulen, Philips tech. Rev. 4, 213-222, 1939.

the walls of the cabinet. This has been avoided by covering the interior of both the high and the low-note cabinets with a sound-absorbent material.

No special problems were involved in this procedure for the high-note cabinet, since there are several materials commercially available that adequately absorb the higher audio frequencies. For the lower frequencies, however, the solution was not so simple. In this case the principle of "panel resonance" was adopted. The inside of the cabinet was fitted with panels mounted on a framework of laths, leaving an air cushion between these panels and the walls of the cabinet (fig. 13). Owing to the

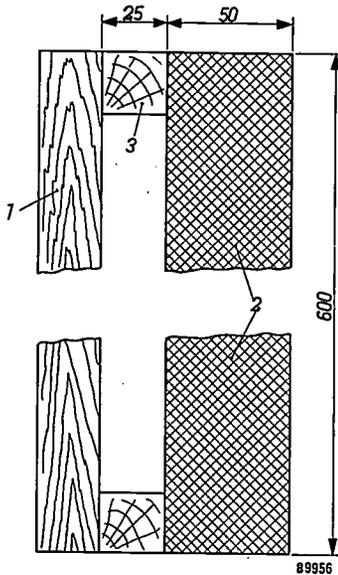


Fig. 13. Standing waves have been avoided in the low-note cabinet by fitting the inside of the walls 1 with panels 2 of absorbent material. The spacing laths 3 provide for an air cushion between wall and panel. Dimensions in mm.

resonator interaction of the mass of the panel with the stiffness of the air cushion, an adequate absorption through the low-note range is achieved, mainly as a result of the dissipation in the clamped-in edges of the vibrating panels. The panels, consisting of an absorbent material, furthermore bring about ordinary absorption in the upper part of the low-note range.

There must be sufficient absorption over a certain frequency range. This range is determined by the following. The most important of the natural vibrations of the cabinet are those for which  $\lambda/4$  or  $\lambda/2$  is equal to one of the interior dimensions of the cabinet. The largest interior dimension is 70 cm, so that the lowest frequency of the natural vibrations is  $c/(4 \times 0.70) = 120$  c/s. By giving the panels a natural frequency  $f_p = 240$  c/s and making this resonance not very selective, adequate absorption is obtained in the entire range from 120 to 420 c/s (in the upper part of this range, assisted by the absorption of the material itself). In connection with dimensioning the

resonator such that  $f_p = 240$  c/s, the following should be pointed out. By analogy with eq. (3), we may write:

$$f_p = \frac{1}{2\pi} \sqrt{\frac{S_a}{M}} \dots \dots \dots (13)$$

$S_a$  represents mainly the stiffness  $S_a$  of the air cushion behind the panel,  $M$  mainly the mass  $M_p$  of the panel; the stiffness of the panel at the edges and the mass of the air moving with the panel may be neglected to a first approximation. According to (7),  $S_a$  can be written as

$$S_a = \frac{\rho_0 c^2 A_p^2}{A_p d_a}, \dots \dots \dots (14)$$

$\rho_0$  being the density of the air,  $A_p$  the area of the panel and  $d_a$  the thickness of the air cushion (panel-wall distance).  $M_p$  can be written as

$$M_p = A_p d_p \rho_p, \dots \dots \dots (15)$$

$d_p$  being the thickness of the panel and  $\rho_p$  the density of the absorbent material. By substituting (14) and (15) in (13) we arrive at:

$$f_p = \frac{c}{2\pi} \sqrt{\frac{\rho_0}{\rho_p d_p d_a}}$$

Now  $c = 340$  m/s and  $\rho_0 = 1.3$  kg/m<sup>3</sup>. The density  $\rho_p$  of the absorbent material used is 55 kg/m<sup>3</sup>. For  $d_p = 50$  mm and  $d_a = 25$  mm (fig. 13) we find  $f_p =$  approx. 240 c/s.

**Loudspeakers with constant impedance**

The self-induction of the speech coil causes its impedance to increase with the frequency (see the dotted curve in fig. 14). If the coil is fed with a frequency-independent voltage, the current, and consequently the force driving the cone, will decrease with increasing frequency. As mentioned in the description of the installation, the loudspeakers used have a virtually constant impedance. This is owing to the introduction of a copper ring ( $K$ , fig. 4) within the coil. This ring operates as a short-circuited winding and thus mainly eliminates the self-induction of the coil<sup>10</sup>). The result is the

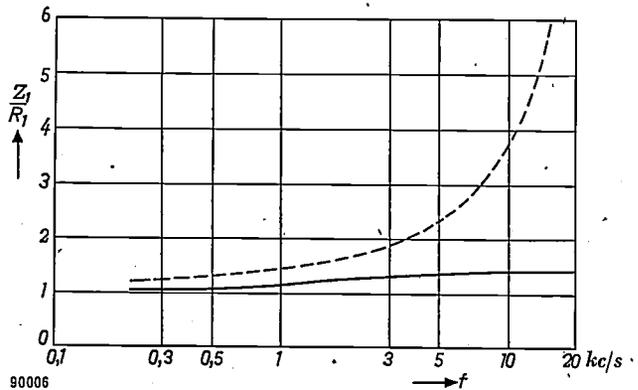


Fig. 14. Impedance  $Z_1$  of a loudspeaker divided by the D.C. resistance  $R_1$ , as a function of the frequency  $f$ . Ordinary loudspeaker: dotted curve; loudspeaker with short-circuiting ring: full curve.

<sup>10</sup>) A ring of this type was mentioned in the Philips tech. Rev. 4, 301, 1939, footnote 1).

impedance curve shown as the fully drawn line in fig. 14. Between the frequencies 400 c/s and 18 kc/s the impedance changes only in the ratio 1 : 1.5; without short-circuiting ring this ratio would be between 1 : 5 and 1 : 6.

The optimum shape and dimensions of the short-circuiting ring can be determined from the following analysis.

Let  $E_1$  be the terminal voltage and  $I_1$  the current through the coil (of resistance  $R_1$  and self-inductance  $L_1$ ), and  $I_2$  the current induced in the short-circuiting ring (of resistance  $R_2$  and self-inductance  $L_2$ ); the impedance  $Z_1 = E_1/I_1$  of the coil is then given by:

$$Z_1 = R_1 + \frac{k^2 L_1 R_2 / L_2}{1 + \left(\frac{R_2}{\omega L_2}\right)^2} + j\omega L_1 \left[ 1 - \frac{k^2}{1 + \left(\frac{R_2}{\omega L_2}\right)^2} \right] + Z' \quad (16)$$

In this equation  $k$  represents the coupling factor  $= \sqrt{M_k^2 / L_1 L_2}$  ( $M_k$  being the mutual induction between coil and ring), and  $Z'$  represents a term accounting for the e.m.f. induced in the moving coil. The motion of the coil, however, does not alter the alternating flux linked by the ring. In order to determine the influence of the ring on the impedance of the coil it is therefore permissible to consider the coil as being stationary, i.e. to put  $Z' = 0$ ; to distinguish  $Z_1$  in this case we shall call it  $Z_{10}$ . As regards "low" frequencies ( $\omega \ll R_2/L_2$ ), equation (16) can be written as

$$Z_{10} \approx R_1 + j\omega L_1, \dots \dots \dots (16a)$$

and as regards "high" frequencies ( $\omega \gg R_2/L_2$ ) as

$$Z_{10} \approx (R_1 + k^2 L_1 R_2 / L_2) + j(1 - k^2) \omega L_1 \quad (16b)$$

The dimensions of the ring should be such that the following conditions are satisfied:

- 1) The frequency at which  $\omega L_2 = R_2$  must be so low that  $\omega L_1$  is still small with respect to  $R_1$  (e.g. below 2 kc/s), so that at "low" frequencies (see eq. (16a))  $Z_{10} \approx R_1$ .
- 2) The coupling between ring and coil must be very tight, so that the factor  $(1 - k^2)$  in (16b) is sufficiently small to keep the term  $(1 - k^2)\omega L_1$  small with respect to the resistance term up to the highest audio frequencies.

Both requirements can be satisfactorily met. It should be noted that compliance with 1) means that also the quantity  $k^2 L_1 R_2 / L_2$  by which the resistance term in (16b) differs from  $R_1$ , is small with respect to  $R_1$ . With an appropriately dimensioned ring, therefore, the efficiency of the loudspeaker is not appreciably reduced.

The question arises whether, with a loudspeaker without copper ring, the two soft-iron boundaries of the air-gap would not act as short-circuiting rings. This effect, however, is negligible, owing to the skin effect, which is far more pronounced in iron, with its fairly high permeability and rather poor conductivity, than in copper. With the copper ring no skin effect is noticeable since the penetration depth of the current is greater than the thickness of the ring, even at 20 kc/s. This means that the current is virtually uniformly distributed over the cross-section of the ring, so that  $R_2$  is practically equal to the D.C. resistance of the ring.

### The amplifier

As mentioned above, the bass-note loudspeakers in their cabinet show a mild resonance peak around

60 c/s. To provide a further damping of this resonance, the amplifier should have a low internal resistance.

The influence of the internal resistance upon the damping can be explained as follows. Owing to the motion of the coil (wire length  $l$ , velocity  $v$ ) an e.m.f.  $Blv$  is induced in it. This e.m.f. operates on the resistance  $R_1$  of the coil in series with the internal resistance  $R_i$  of the amplifier ( $R_i$  measured at the output terminals). These two resistances thus have to dissipate a power  $(Blv)^2 / (R_1 + R_i)$ . Instead of this, let us imagine an equal dissipation in an imaginary mechanical resistance  $R_m'$ :

$$\frac{(Blv)^2}{R_1 + R_i} = R_m' v^2,$$

and hence

$$R_m' = \frac{B^2 l^2}{R_1 + R_i}$$

The mechanical damping resistance already present is consequently raised by this amount,  $R_m'$  being larger the lower the internal resistance  $R_i$  of the amplifier.

An effective means of reducing the internal resistance is the use of voltage feed-back, i.e. part of the output voltage is applied in anti-phase to the input voltage of the output stage or of a previous stage<sup>11)</sup>.

The splitting-up into two frequency ranges after the amplifier requires an amplifier that is adequately designed to prevent any appreciable intermodulation. The amplifiers AG 9000 and AG 9006 (fig. 1) are very suitable in this respect, respectively supplying 10 and 20 W output at 2% intermodulation. The former amplifier has been designed for loudspeakers with the conventional low resistance (two 7 Ω speakers in series) and is therefore equipped with a step-down output transformer. The amplifier AG 9006, on the other hand, contains a novel output circuit, capable of directly feeding high-impedance loudspeakers<sup>12)</sup>; the output transformer with its inevitable distortion is thus obviated here.

<sup>11)</sup> B. D. H. Tellegen, Philips tech. Rev. 2, 289-294, 1937.

<sup>12)</sup> An article on output circuits for high-impedance loudspeakers will appear shortly in this Review.

**Summary.** The high-fidelity loudspeaker installation for the home described here comprises a corner cabinet with two bass-note loudspeakers, two separate boxes for the higher frequencies, each containing a double-cone loudspeaker, and filters for splitting up the audio spectrum into two ranges; one below and one above 420 c/s. By appropriately positioning the high-note loudspeakers, good diffusion of the sound can be achieved. The listener then hears mainly indirect sound, just as in the concert hall. The so-called hole-in-the-wall effect, which is a drawback of reproduction by a single loudspeaker, is thus eliminated.

One cause of non-linearity in moving-coil loudspeakers is the amplitude-dependence of the stiffness of the cone suspension. The influence of this has been greatly reduced, 1) by doubling the number of loudspeakers, so that the cones can vibrate with a smaller amplitude, and 2) by adding a considerable, practically linear air stiffness. The latter was achieved by using completely closed cabinets. For low-note reproduction this has the additional advantage that the air waves cannot travel around the cone, whilst the high-note speakers will not emit any sound to the rear. Standing waves inside the cabinet are prevented by lining the cabinets with an absorbent material (in the low-note cabinet in the

form of panels separated from the walls by an air cushion). Non-harmonic distortions (intermodulation and Doppler effect) have been greatly reduced by dividing the total frequency range into two parts. The overall reproduction extends from about 20 c/s to about 18 kc/s, the high upper limit resulting from the use of double-cone loudspeakers. Inside the coil of the loudspeakers is a fixed short-circuited ring, which virtually eliminates the self-inductance of the coil, so that the impedance rises only very little with the frequency. The amplifier feeding the installation should have a low internal resistance, which can be achieved by voltage feedback. Amplifiers type AG 9000 and AG 9006 are very suitable in this respect.

## ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk \* can be obtained free of charge upon application to the Philips Research Laboratory, Eindhoven, Netherlands.

- 2393:** H. B. G. Casimir: Anticipations . . . a continental view of physics and its future (Physics Today 9, April 1956, pp. 13-16).

General remarks on the state of fundamental physical research in Europe and elsewhere. The author discusses some of the conditions under which research is pursued, its changing background and organization, the problem of language and the present shortage of theoretical physicists in Europe. The article concludes with some remarks on the aims and responsibilities of physicists.

- 2394:** P. C. van der Willigen and L. F. Defize: CO<sub>2</sub>-booglassen van staal (Lastechniek 22, 69-77, 1956). (CO<sub>2</sub> arc welding of steel; in Dutch).

Arc welding in a CO<sub>2</sub> atmosphere can be used on various types of mild steel, including (according to the investigations reported here) rimmed steel, provided that the consumable electrode contains sufficient de-oxidizing elements. The remarkable transfer of the droplets across the arc has been studied using a high-speed camera. Because the whole CO<sub>2</sub> welding process is chemically relatively simple, quantitative measurements can be made from which a good picture is obtained of the various reactions. It was found that welding in CO<sub>2</sub> is about "equivalent" to welding in argon + 9% oxygen, with the exception of the reaction which takes place with carbon. During welding, about 2½% CO does not react with the oxygen from the air (calculated on the CO<sub>2</sub> being used). The main features of CO<sub>2</sub> arc welding are: absence of hydrogen, low nitrogen content of the weld, less trouble with the arc length in comparison with argon, very deep penetration which shows a U-form. Because of the latter feature unbevelled butt welds can be made with low heat

input and low wire consumption. CO<sub>2</sub> arc welding can be performed more easily by machine welding than by hand welding, mainly because of the remarkable metal transfer, the very short arc and the high travel speeds.

- 2395:** H. Rinia: Optics in television (Problems in contemporary optics, Istituto Nazionale di Ottica, Arcetri-Firenze, 1956, pp. 492-507).

Television, which is primarily an electronic technique, has caused a renewed interest in optical problems. It has led to a demand for certain optical systems: Schmidt systems, variable-focus lenses and dichroic mirrors are examples. Television optics has introduced a new stimulus into the already existing question as to how to judge the performance of an optical system. Finally television techniques may provide new tools for optics.

- 2396:** J. L. Meijering and H. K. Hardy: Closed miscibility gaps in ternary and quaternary regular alloy solutions (Acta metallurgica 4, 249-256, 1956).

Although the regular-solution model is only an approximation, it may yield useful semi-quantitative results in predicting the form of ternary miscibility gaps from binary data. The spinodal equation of quaternary regular solutions has been derived and the coordinates of second-order critical points are given as a function of the six binary interaction parameters. Quaternary critical points are of special interest, as they are connected with closed miscibility gaps. The conditions which cause a ternary critical temperature to be raised by the addition of a fourth component are given. The results are applied to miscibility gaps in liquid and f.c.c. alloys.

**2397:** J. L. Meijering: On thermodynamical analysis by calorimetry alone (*Acta metallurgica* 4, 333-335, 1956).

Letter in which the author discusses and supports the plea of Oelsen, Schürmann and Heynert for thermodynamical analysis by purely calorimetric methods. Some extensions of this idea are put forward.

**R 300:** P. Zalm: The electroluminescence of ZnS type phosphors (*Philips Res. Rep.* 11, 417-451, 1956).

Continuation of **R 298**. In section 4 of this paper it is shown that the voltage dependence of the emittance is due to the presence of a Mott-Schottky barrier and may be expressed by  $H = H_0 \exp(-cV^{-1/2})$ . A theoretical model is drawn up for crystals embedded in a dielectric. This is tested against a number of experimental results reported in section 5 of the paper. Finally a study is made of the energy efficiency of electroluminescent phosphors.

**R 301:** J. Volger and J. M. Stevels: Further experimental investigation of the dielectric losses of various glasses at low temperatures (*Philips Res. Rep.* 11, 452-470, 1956).

The results of loss measurements at low temperatures on a number of glasses are given. These are discussed qualitatively in connection with the glass structure. The importance of impurities in fused silica is shown. The influence of thermal treatment is also considered. Contrary to experience with quartz crystals, no effect of irradiation (formation of colour centres) upon the dielectric losses at low temperatures is found. A comparative survey of the various loss mechanisms found hitherto in the vitreous and crystalline states is given.

**R 302:** W. Ch. van Geel and C. A. Pistorius: On the residual voltage with electrolytic capacitors (*Philips Res. Rep.* 11, 471-478, 1956).

After an electrolytic capacitor, e.g., Al/Al<sub>2</sub>O<sub>3</sub>/electrolyte, has been charged and discharged, a voltage is found to develop anew across the capacitor plates. This is called the residual voltage, and it is found to be proportional to the product of the field strength in the oxide layer and the thickness of this layer. It is assumed that Al<sup>3+</sup> ions in the lattice have moved to adjacent positions, considering the fact that there is a large number of vacancies in the Al<sub>2</sub>O<sub>3</sub> lattice. A simple computation shows that for an average shift of an Al<sup>3+</sup> ion over a distance of 2 Å, 0.1% of the available Al<sup>3+</sup> ions have been displaced.

**R 303:** C. Z. van Doorn and Y. Haven: Absorption and luminescence of colour centres in KCl and NaCl (*Philips Res. Rep.* 11, 479-488, 1956).

Crystals of KCl and NaCl, coloured additively and quenched from high temperature to liquid-nitrogen temperature, show the *F*-absorption band and when irradiated in this band an emission in the infrared. If kept for a short time at room temperature a new absorption band (*M*-band) develops and irradiation in this band gives rise to a new emission band. If an *M*-absorption band is present, this new emission band is also found when the crystal is irradiated in the *F*-band. An explanation for this is offered. After long periods of ageing, *R*-bands also develop; irradiation in the *R*-bands as well as in the *F*-band gives rise to a third emission band.

**R 304:** A. I. Luteijn and K. J. de Vos: Permanent magnets with (*BH*)<sub>max</sub> values over ten million gauss oersteds (*Philips Res. Rep.* 11, 489-490, 1956).

Brief note announcing the attainment of (*BH*)<sub>max</sub> values up to 11 million gauss oersteds at a remanence of 11800 gauss and a coercivity of 1315 oersteds in permanent magnets of "Ticonal" X (titanium-containing), by preparation in an atmosphere of pure argon, using very pure raw materials. Rods of these alloys produced by pulling from the melt contain large crystals with a [100] axis nearly parallel to the direction of pulling, and a subsequent heat treatment in a magnetic field produces the high values reported above.

**R 305:** J. D. Fast: The influence of impurities on the recrystallization texture of cold-rolled 3% silicon iron (*Philips Res. Rep.* 11, 490-491, 1956).

In the preparation of grain-oriented silicon-iron sheet, whereby a preferred direction of magnetization is obtained by combined rolling and heat-treatment, the results are often poorly reproducible. The present note refers to an investigation which shows that the grain-orientation effect is brought about by impurities: the effect cannot be obtained in very pure silicon iron. It has been found that, in particular, small amounts of nitrogen introduced into the alloy before the cold-rolling produce a very good and reproducible texture. To achieve the optimum magnetic properties, the nitrogen must be expelled during a final heat-treatment.

# Philips Technical Review

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

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

## THE EC 57, A DISC-SEAL MICROWAVE TRIODE WITH L CATHODE

by G. DIEMER, K. RODENHUIS and J. G. van WIJNGAARDEN.

621.385.3.029.6:621.3.032.213.2

---

*The triode is the cornerstone of the still rapidly developing radio industry. Many tubes have been derived from the triode, which itself continues to play an important role. It is a remarkable fact that the triode has also won a place for itself in the new branches of very-high-frequency radio engineering, for which it was at one time thought that only velocity-modulated tubes were suitable. The designers of microwave triodes have, of course, made good use of the latest discoveries. The present article deals, among other things, with the importance in this respect of the introduction of the L cathode.*

---

### Introduction

It was at one time thought that the triode, owing to transit-time effects and the low impedances of its inter-electrode capacitances, would not be able to compete in the microwave range with tubes such as magnetrons and klystrons in which transit-time effects are utilized, and travelling-wave tubes, in which, moreover, concentrated capacitances are avoided. Many tube designers have endeavoured, however, to overcome the triode's limitations by still further reducing its dimensions. As a result of combining the work done in different spheres (the manufacture of extremely thin wire for the grids, better cathodes and metal-glass seals, and the refinement of assembly methods), these endeavours have been rewarded with success. For many purposes, triodes are now well able to compete with other tubes in the centimetre wave range, e.g. as oscillators of low power (a few mW), as low-noise wide-band amplifiers and as oscillators and amplifiers of moderate power output (a few watts). The introduction of the L cathode has opened up new perspectives for high-power triodes (20 W or more), the development of which has not yet been concluded.

For certain applications, triodes offer considerable advantages over tubes of the klystron and travelling-wave type. Because of their low supply

voltages and the absence of focusing devices, equipment employing triodes can be simplified, reduced in weight and, in many cases, made more cheaply. Moreover, triodes have a large bandwidth, which is an advantage most klystrons lack.

As an example of a microwave triode we shall discuss in this article the type EC 57, which is shown in *fig. 1*. This tube was developed mainly for use in link transmitters, but it can also be put to good use in small, mobile radar units. Up to 4000 Mc/s the tube delivers a power of at least 1.8 W as an amplifier (for 16 Mc/s bandwidth) and 3 W as an oscillator. In designing the EC 57 we have been aided by the insight gained in recent years into the operation of microwave triodes. The L cathode<sup>1)</sup> has also been an important element in the design.

### New considerations on the operation of microwave triodes

The operation of triodes at very high frequencies has latterly come to be much better understood.

---

<sup>1)</sup> H. J. Lemmens, M. J. Jansen and R. Loosjes, Philips tech. Rev. 11, 341-350, 1949/50.

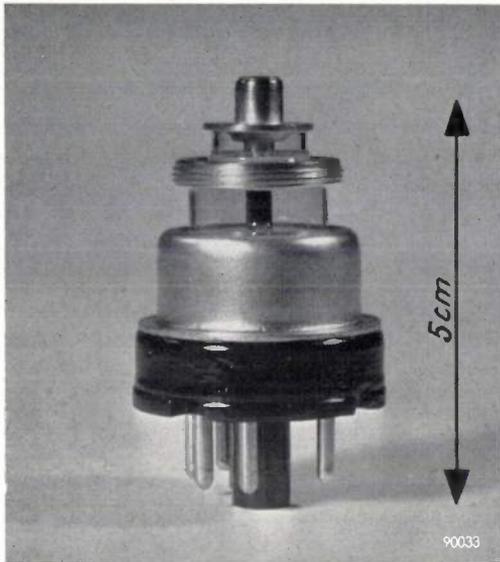


Fig. 1. The microwave triode EC 57.

It was known even before 1940 that, in order to combat transit-time effects, it is necessary not only to make the clearances between the electrodes extremely small but also to use the densest possible electron current. It may be derived that the transit time  $\tau$  of the electrons travelling from cathode to grid is approximately given by

$$\tau = 6.7 \times 10^{-10} \left( \frac{d}{j} \right)^{\frac{1}{2}} \text{ sec.} \quad (1)$$

in which  $d$  is the distance in cm from cathode to grid and  $j$  is the current density in A/cm<sup>2</sup>. The favourable influence of a large current density may be broadly understood by the fact that the higher potential needed in the grid plane in order to obtain a denser current gives rise to higher electron velocities and thus to shorter transit times. For a closer derivation of (1) we must consider the potential distribution and the movement of the electrons in the space between the cathode and grid of a planar triode<sup>2)</sup>. The triode operates under space-charge conditions, that is to say only some of the electrons emitted by the cathode reach the grid plane. The potential distribution between cathode and grid in this case is set out in *fig. 2* (the grid being represented as the anode of a diode). Near to the cathode there is a minimum in the potential curve. For the derivation of (1) it is assumed that the electrons start in this potential minimum (the quantities  $d$  and  $j$  being defined accordingly).

<sup>2)</sup> See e.g. G. Diemer and K. S. Knol, Philips tech. Rev. 14, 153-164, 1952/53.

In reality, the result of a potential distribution as in *fig. 2* is that only those electrons that leave the cathode with a sufficiently high initial velocity are able to pass the potential minimum and travel to the grid; the other electrons reverse their direction before the potential minimum and return to the cathode — a subject which will be referred to later.

Another important point affecting the operation of a microwave triode is that the resistance and self-inductance of the supply leads must be kept as low as possible<sup>3)</sup>. In particular the self-inductance of the grid lead must be kept low in order to avoid unwanted feedback. These principles have been put into effect in the microwave triodes by making the inter-electrode spacings extremely small and by using disc-seals<sup>4)</sup>. The structure of a disc-seal triode is shown diagrammatically in *fig. 3*. The disc-seals form an effective screen between the input and output circuits and may conveniently form part of the wall of a cavity resonator.

The disc-seal triodes described in the book quoted under<sup>4)</sup> were the first triodes to be used at

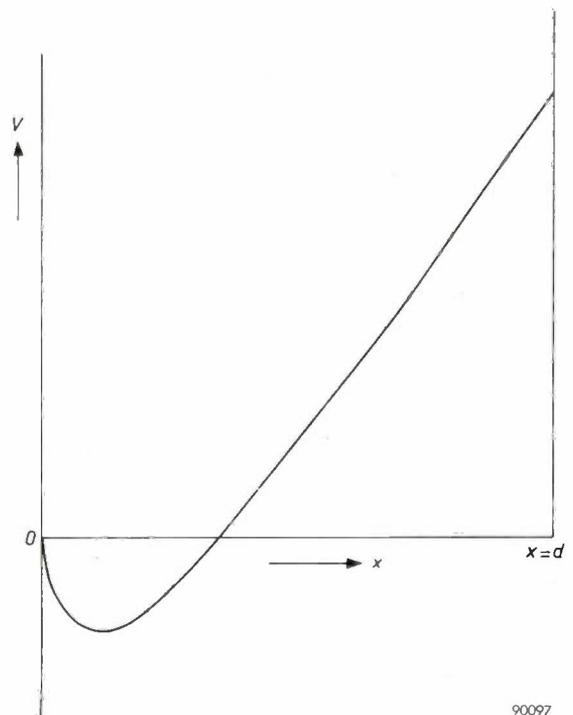


Fig. 2. The potential  $V$  as a function of the distance  $x$  along the cathode ( $x = 0$ ) — anode ( $x = d$ ) space for a diode operating under space-charge conditions.

<sup>3)</sup> See e.g. M. J. O. Strutt and A. van der Ziel, Philips tech. Rev. 3, 103-111, 1938; 5, 172-181 and 357-362, 1940.

<sup>4)</sup> See D. R. Hamilton, J. K. Knipp and J. B. Horner Kuper, Klystrons and microwave triodes, McGraw-Hill, New York 1948, p. 163 et seq.

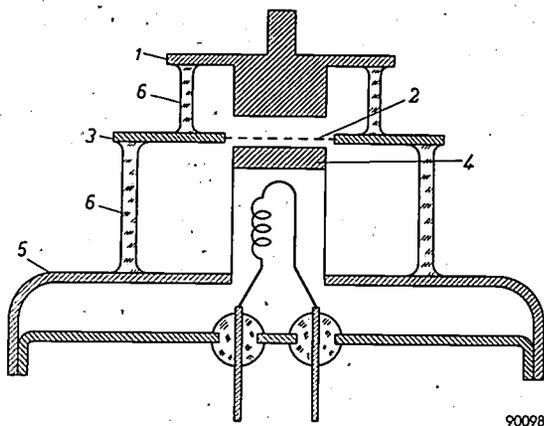
frequencies from 1000 to 2000 Mc/s. In the EC 57, however, which can be operated up to 4000 Mc/s, another factor of primary importance, apart from the self-inductance, is the resistance of the electrode

by the tube is lost by dissipation in  $r$ . (This loss would be much greater in chrome-iron discs without silver plating,  $r$  then being  $0.75 \Omega$ .)

The effect of further increasing the frequency is a decrease in the gain  $G$  (while the bandwidth at optimum matching ( $g = g'$ ) also becomes larger until the value  $G = 1$  is reached at the frequency limit). Because of the low value of  $r$  and the large ratio of transconductance to capacitance obtained with the L cathode (see below), and also because of efficient dimensioning, the gain of the EC 57 does not begin to drop sharply until about 6000 Mc/s, and the frequency limit is about 15000 Mc/s. As regards dimensioning, for the purpose of obtaining the highest possible frequency limit, the useful electrode surfaces should be made so large that the inter-electrode capacitances have about the same value as the unavoidable capacitances of edges, electrode leads etc. <sup>5)</sup>

There are certain other factors that affect the operation of microwave triodes. It has been found, for example, that the high-frequency currents in the grid wires can themselves give rise to feedback <sup>6)</sup>. Moreover, a new type of loss has been discovered, known as total-emission damping <sup>2) 7)</sup>, which is due to the high-frequency energy consumption by the numerous electrons which reverse their direction before the potential minimum and return to the cathode. The effect of this damping is greater the closer the grid is brought to the potential minimum — e.g. by reducing the dimensions. The same electrons also give rise to additional noise, known as total-emission noise.

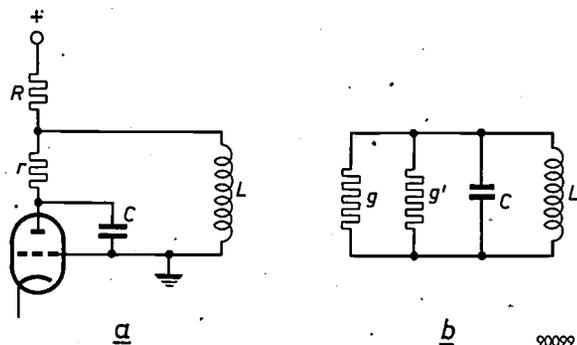
Differences in transit time profoundly affect the operation of the tube. For instance, even at transit times of several cycles a tube would still have a reasonable amplification provided the transit times of all electrons were the same <sup>8)</sup>. It would then be possible to reduce the noise in the output circuit by the use of special circuitry (noise compensation) <sup>9)</sup>. The spread in the transit times results in phase differences between the signal currents induced by the various electrons, one consequence of which is reduced amplification. Another important consequence, however, is that the correlation is lost between the different noise currents, and with it the possibility of noise compensation. The spread in



90098

Fig. 3. The structure of a disc-seal triode (schematic); 1 anode disc with anode, 2 grid, 3 grid disc, 4 cathode, 5 cathode disc, 6 glass rings.

leads (i.e. the discs; in the EC 57 of silver-plated chrome-iron). The influence of this series resistance  $r$ , which is particularly noticeable at the output side of the tube, may be explained with reference to the circuit in *fig. 4a*. Here,  $R$  is the load resistance in which the useful power is dissipated;  $g = 1/R$  (see equivalent circuit *fig. 4b*) is the useful load conductance. The resistance  $r$  of the leads, which,



90099

Fig. 4. a) Grounded-grid triode circuit, b) equivalent circuit.

owing to the skin effect, is proportional to the root of the frequency  $\omega$ , causes, together with the capacitance  $C$  (largely the capacitance of the tube itself) an equivalent parallel conductance

$$g' = \omega^2 C^2 r \propto \omega^{2.5} C^2 \dots \dots (2)$$

At 4000 Mc/s the minimum value to which  $r$  can be reduced is  $0.15 \Omega$ ; this is achieved by plating the discs with silver. With this value of  $r$ ,  $g'$  becomes  $1/10500 \Omega^{-1}$ . If  $R$  is chosen such that the bandwidth is 100 Mc/s, only 17% of the total power delivered

<sup>5)</sup> G. Diemer and K. Rodenhuis, Philips Res. Rep. 7, 36-44, 1952.  
<sup>6)</sup> See G. Diemer, Philips Res. Rep. 5, 423-434, 1950.  
<sup>7)</sup> A. van der Ziel and A. Versnel, Philips Res. Rep. 3, 13-23, 1948.  
<sup>8)</sup> A. van der Ziel, Philips Res. Rep. 1, 381-399, 1946.  
<sup>9)</sup> G. Diemer and K. S. Knol, Philips tech. Rev. 14, 236-244, 1952/53.

transit times has many causes, among which are the finite initial velocity of the electrons, the deviations from the ideal planar triode structure and local variations in work function.

### The role of the cathode

In the development of the EC 57 the L cathode was of fundamental importance. Its use resulted in substantial improvements upon microwave triodes with oxide cathodes. One advantage of the L cathode is that, owing to its metallic surface, it gives rise to fewer high-frequency losses than the oxide cathode, the surface of which is a semi-conductor. Furthermore, the L cathode is less susceptible than the oxide cathode to damage from the ion bombardment that arises during the degassing of the anode in the manufacturing stage. (In the disc-seal triode the degassing process cannot be effected by high-frequency heating, as in other tubes, and therefore is performed by electron bombardment from the cathode, which gives rise to a kind of gas discharge between the electrodes.) However, the most important improvements obtained by using the L cathode are based upon the fact that it can tolerate a much higher continuous current density than the oxide cathode although it has a smaller saturation current density. For example, the L cathode can easily yield  $0.5 \text{ A/cm}^2$  at a saturation of  $2 \text{ A/cm}^2$ . The saturation of the oxide cathode is higher, e.g.  $10 \text{ A/cm}^2$ , but of this only  $0.2 \text{ A/cm}^2$  can be utilized if the cathode is to have a reasonable life.

It has already been pointed out that a higher current density is advantageous for reducing electron transit times — and hence also the spread in transit times. The fact that with the L cathode the ratio of anode current to saturation current is much greater than with the oxide cathode has the important consequence of allowing the distance from the cathode to the potential minimum to be reduced to e.g.  $6 \mu$  with the given numerical values, as compared with  $10 \mu$  in the case of the oxide cathode. The effect of this is to reduce considerably the transit time as well as the number of the electrons that return before the potential minimum. This results in a substantial reduction of the total-emission damping and the total-emission noise.

A further advantage of the higher current density is that a smaller cathode surface is needed for the same transconductance. This means smaller tube capacitances and consequently lower losses in the circuits. Moreover, owing to the correspondingly smaller grid surface, the effect of magnetic feedback through the grid wires is smaller.

Finally, the higher current density of the L cathode as compared with the oxide cathode makes it possible to obtain a higher power output (e.g. 20 W or more) at a given bandwidth. Since the tube described in this article has been developed for moderate power output, we shall not go into this matter here.

Together with the advantages mentioned, the L cathode has certain drawbacks, the effects of which can, however, be compensated by suitable design. The drawbacks referred to are the following:

- 1) Owing to the higher temperature of the L cathode ( $1050^\circ\text{C}$  as against  $750^\circ\text{C}$  for the oxide cathode), the radiation from the cathode makes the grid wires appreciably hotter, so that they are more liable to bend as a result of expansion.
- 2) The higher heater power calls for special care in the construction of the heater.
- 3) The more rapid evaporation of barium and barium oxide involves the risk that these materials will form deposits on the grid wires (thereby altering the characteristics of the tube) and on the glass envelope (resulting in additional losses).

The methods that have been adopted to counter these effects, will be discussed below.

### Design of the EC 57 triode

We shall consider the components and the assembly of the EC 57 in more detail by reference to *fig. 5*, which shows a simplified cross-section of the tube.

#### *The envelope*

The envelope consists of the glass walls (6) fused to the chrome-iron components (1), (3) and (5). The latter carry the electrodes and are conveniently termed discs, although their shape is in fact more complicated.

We have already shown that the resistance of the discs must be as small as possible. For this reason the chrome-iron is silver-plated. The grid disc is so designed as to make it impossible for barium evaporating from the cathode to move in a straight line to the glass ring between grid and anode. The barium is thus intercepted at a place where it can do no harm. For the same purpose a shielding ring (9) is incorporated which keeps the glass wall between grid and cathode clean.

In some disc-seal triodes the discs are made of copper. Copper being relatively soft, it is then necessary to use ceramic spacers inside the envelope (stacking construction) in order to fix the electrodes

at the required small clearances. The disadvantage of these spacers is that they introduce extra capacitance in the tube. Moreover, because they are usually at a fairly high temperature they give rise to rather large dielectric losses. In the EC 57 using only the chrome-iron discs and glass walls, it proved possible by means of a special method of

the activation temperature of the cathode. The cathode is brought to the appropriate temperature by the heating of a filament, which is sintered in the molybdenum cathode tube with the aid of aluminium oxide, thus ensuring good heat transmission. In this way the temperature of the filament does not become so high as to shorten its useful life.

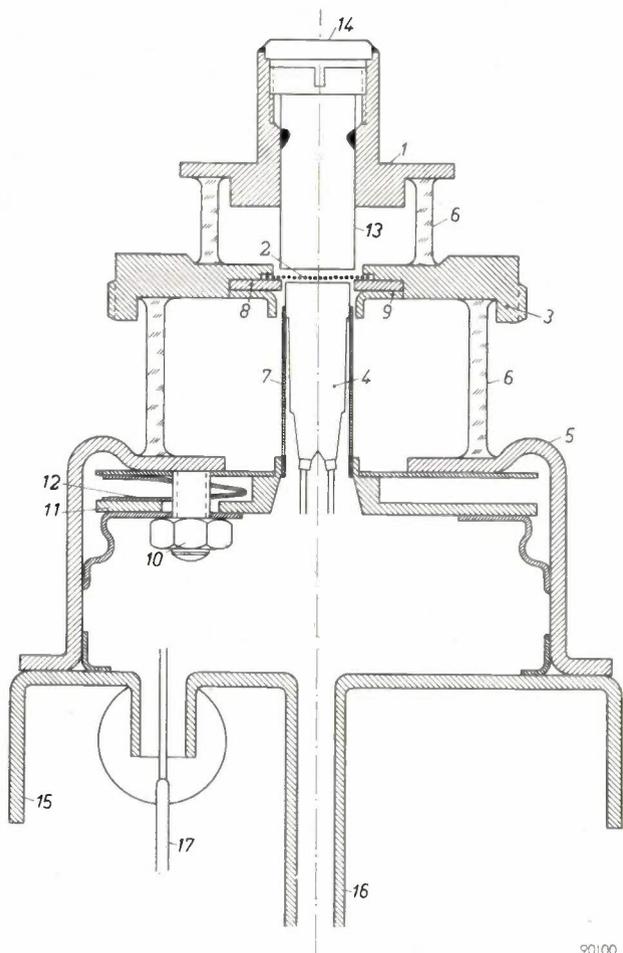


Fig. 5. Simplified section of the EC 57, showing most important parts: 1 anode holder, 2 grid, 3 grid disc, 4 cathode, 5 cathode disc, 6 glass rings, 7 cathode holder, 8 molybdenum frame, 9 shielding ring, 10 nut, 11 cathode base, 12 contact spring, 13 anode, 14 anode cap, 15 base plate, 16 exhaust stem, 17 filament connecting pin.

adjustment (which will be discussed below) to obtain the required grid-cathode separation of  $40 \mu$  (and *a fortiori* the much greater grid-anode separation). This considerably reduces the capacitances and the losses.

#### The cathode

The L cathode (4) is fixed to the cathode base (11) by means of tantalum foil in tubular form (7) (thickness  $0.012 \text{ mm}$ ). Tantalum is easy to work, its heat conductivity is not too high, its electrical conductivity is tolerable and it remains sturdy at

#### The grid

Tungsten wire (diameter  $7.5 \mu$ )<sup>10</sup> is used for the grid and is wound on a molybdenum frame (8) with a  $4 \text{ mm}$  square aperture; since the pitch is  $48 \mu$  there are more than 80 turns in the frame. Fig. 6 shows a photograph of this grid enlarged 9 diameters.

During operation the grid wires, as we have seen, become hot as a result of radiation from cathode and anode. In fact, however, a high temperature has a beneficial effect in that it considerably reduces the formation of barium deposits on the grid wires. The thinner the wires the higher their temperature, owing to the diminished heat conductivity to the grid disc. The diameter of the wires is chosen such that, at the temperature in question, the barium evaporates from the wires almost as fast as it arrives from the cathode. If the temperature of the grid wires is too high there is admittedly a great risk of grid emission. In the EC 57, however, this unwanted emission is adequately suppressed by covering the grid wires with a thin layer of gold.

If no precautions were taken the wires would be deformed by expansion. For this reason they are

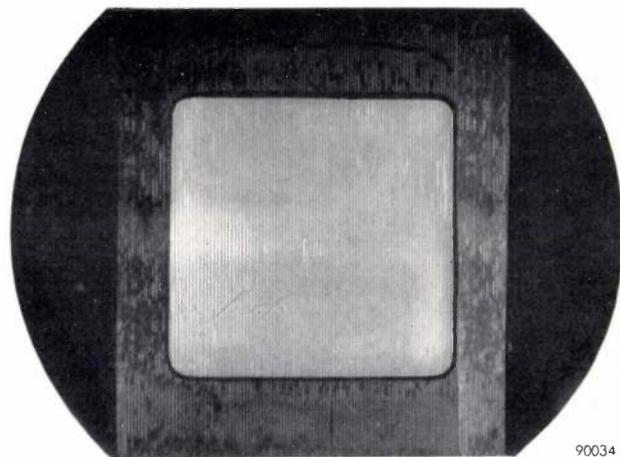


Fig. 6. Photograph of grid enlarged  $9\times$ .

<sup>10</sup> For the manufacture of thin tungsten wire, see: L. Schultink and P. G. van Zanten, Philips tech. Rev. 18, 222-228, 1956/57.

mechanically pre-stressed. By braking the coil of wire during the winding process, the wire is stressed to about 60% of its tensile strength.

After winding, the turns are gold-soldered to the frame and then removed from one side. The soldering reduces the tensile strain in the grid wires. The reasons for this are many and complicated, the most important being the fact that the molybdenum frame has a somewhat greater coefficient of expansion than the tungsten wires, and that the yield point of the wires is lowered at the soldering temperature (about 1200 °C). A tungsten wire of 100  $\mu$  diameter is used as a well-defined wedge and is pressed into the frame perpendicular to the grid wires, on the side free of wires; by this means the frame is made to expand by a precise amount (a few microns), so that the strain returns to the required value. Since the frame aperture is square, the stress in all the wires is increased by the same extent. The tensile stress in the wires is then measured as follows. The grid is placed under a microscope and a flat metal plate is mounted a fraction of a millimetre away, parallel to the grid wires. A high alternating voltage is then applied between grid and plate. At a certain frequency the wires begin to resonate, and the resonance frequency (about 20 000 c/s) is related simply to the tensile stress in the wire. In this way it is possible to measure the stress of all wires separately. With tungsten wire of thickness 7.5  $\mu$ , the required tensile stress in the wires amounts to about 10 gr per wire. The grid can then tolerate, without deformation, a dissipation of about 1 W, which is amply sufficient for efficient operation.

#### Assembly of the components

After the grid (8) (fig. 5) has been placed in position, the cathode is fitted in the cathode ring (5) and temporarily fixed with three nuts (10). Between the cathode base (11) and the cathode ring there is a 3-lipped molybdenum spring (12) which presses the cathode base against the nuts, thus allowing accurate alignment. The distance between grid and cathode is adjusted under a microscope, the optical axis of which coincides with the tube axis, so that it is possible to look through the aperture of the anode holder (1). A narrow, parallel beam of light (see fig. 7) is obliquely projected on to the cathode: line shadows of the grid wires are then seen on the cathode. By making the shadow of each wire at a given angle of incidence fall exactly under the adjacent wire, the desired spacing can be very accurately adjusted. To ensure that the shadows

do not fall under the wrong wires, a capacitance measurement is also carried out, which serves for the rough adjustment. Perfect parallelism is obtained if the shadows over the whole surface lie centrally under the wires. This having been done, the cathode base is fixed by welding.

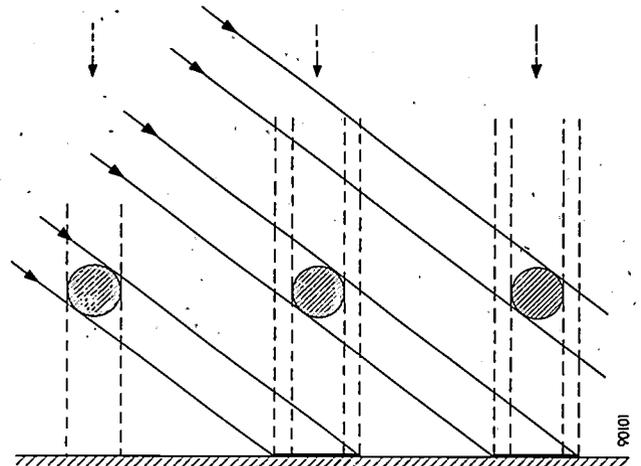


Fig. 7. Precision adjustment of spacing between grid and cathode by shadow method.

The anode (13) is then screwed into the anode holder (1), and again, by means of a capacitance measurement, adjusted to the correct distance from the grid. The anode cap (14) is welded to the anode holder to produce a vacuum-tight seal. Finally, the base plate (15) with its exhaust stem (16) is welded vacuum-tight to the cathode ring. The tube can now be pumped and outgassed.

#### Tube data

From the tube data set out in Table I it will be seen that two values are given for the capacitances between grid and cathode and between grid and anode, viz.: the static values  $C_{gk}$  and  $C_{ag}$  and the high-frequency values  $C_{gk}'$  and  $C_{ag}'$ . The latter values are smaller than the static values owing to the fact that at microwave frequencies only part of the electrode leads acts as a capacitive load.

Table I. Principal data of the EC 57.

Anode voltage	180 V
Anode current	60 mA
Transconductance	20 mA/V
Amplification factor	43
Static grid-cathode capacitance $C_{gk}$	3.3 pF
H.F. grid-cathode capacitance $C_{gk}'$	1.8 pF
Static grid-anode capacitance $C_{ag}$	1.6 pF
H.F. grid-anode capacitance $C_{ag}'$	0.6 pF
Cathode-anode capacitance $C_{ak}$	0.04 pF
Life	about 5000 h

The life of the tube is limited mainly by the barium evaporation, in spite of the countermeasures adopted. It has been found, however, that barium evaporation can be appreciably reduced by small modifications in the cathode, so that further improvements in the life of the tube may be expected.

#### Microwave impedances of the tube

The impedances at microwaves can be calculated with the aid of theoretical considerations based on ideal triodes<sup>11)</sup>. Actual triodes differ so much from the ideal that the theory can provide no more than a rough approximation; in practice, the impedances can be determined accurately only by experiment. However, the theory is useful in as much as it shows how the transconductance, for example, depends upon a number of fundamental quantities. Thus, the transconductance at high frequencies is shown to fall only slowly with increasing transit-time angles. The transit-time angle  $\theta_1$  between cathode and grid is, by definition, the average transit time of the electrons over this distance, multiplied by the angular frequency  $\omega$  of the high-frequency signal. The transit-time angle  $\theta_2$  between grid and anode is defined analogously. The relevant values for the EC 57 at 4000 Mc/s are:

$$\theta_1 = 3.36 \text{ radians,}$$

$$\theta_2 = 1.46 \text{ radians.}$$

At these values the theoretical transconductance has fallen to 75% of that at low frequency.

The high-frequency input conductance is measured by determining the quality factor at 4000 Mc/s of a cavity resonator mounted between cathode and grid. The grid-anode capacitance in this case must be short-circuited to the high-frequency currents; this can be done by enclosing this capacitance in a second cavity resonator with a resonance frequency differing widely from that of the first. In this way we found a total input conductance of 15 mA/V due to the circuit and the electrons together. By measuring the damping of the cut-off tube, we were able to separate the electronic damping from the other losses (losses in the circuit and in the glass parts of the tube). The latter losses can be represented by a resistance shunted across the grid-cathode capacitance. This resistance amounts to 1700  $\Omega$  at 4000 Mc/s, corresponding to a cut-off conductance of 0.6 mA/V, so that the electronic

input damping, corresponding to the transconductance, comes to somewhat more than 14 mA/V. This value of the transconductance, which is thus large with respect to the cut-off conductance, is about 30% smaller than the transconductance at low frequency (20 mA/V), in good agreement with the theory. It also appears that there is little influence from total-emission damping in the normal operating region (the above value was measured at an anode current of 30 to 50 mA).

It has already been stated that the cut-off conductance  $g'$  of the anode-grid space (equation 2) corresponds to a parallel resistance of 10.5 k $\Omega$ , and that this damping appears to have little effect at a bandwidth of 100 Mc/s. At the transit-time angles mentioned, the effect of the electronic output conductance which has a negative value, is also very slight (see<sup>11)</sup>).

To investigate the feedback from the output to the input, measurements were made on an amplifier designed for operation with the EC 57<sup>12)</sup>. The change in the input impedance was determined when the anode load was varied. The feedback was found to be caused by:

- 1) the effect of the anode voltage on the electron current;
- 2) the capacitance  $C_{ak}$ ;
- 3) the high-frequency currents flowing in the grid wires<sup>6)</sup>.

The feedback depends on the distance between grid and anode, a parameter that can be selected within fairly wide limits. At a certain distance between grid and anode the feedback was reduced to zero. Since, however, the gain at the same bandwidth was found to be higher with slight feedback than without<sup>13)</sup>, the grid-anode spacing was adjusted accordingly.

#### Microwave amplification and noise

For microwave amplification the triode will always be used in a grounded-grid circuit<sup>8)9)</sup>, in which, owing to the high electronic damping of the input circuit, the bandwidth of the amplifier is almost entirely determined by the bandwidth of the output circuit. Neglecting the feedback we may express the product of the power amplification  $G$  and the bandwidth  $B$  by the following formula<sup>14)</sup>,

<sup>12)</sup> This amplifier, the circuitry and coupling elements of which called for special design, will be dealt with in an article shortly to be published in this Review.

<sup>13)</sup> For a discussion of this phenomenon see A. van der Ziel and K. S. Knol, Philips Res. Rep. 4, 168-178, 1949.

<sup>14)</sup> See e.g. A. H. W. Beck, Thermionic valves, Univ. Press Cambridge 1953, p. 419.

<sup>11)</sup> F. B. Llewellyn and L. C. Peterson, Proc. Inst. Rad. Engrs. 32, 144-166, 1944.

provided the internal resistance of the tube is large with respect to the anode load:

$$GB = \frac{S^2}{2\pi g_i C_{ag}'},$$

where  $S$  is the absolute value of the high-frequency transconductance,  $g_i$  the total input conductance and  $C_{ag}'$  the high-frequency capacitance of the anode circuit, as already defined<sup>14</sup>). Filling in the values determined, we find

$$GB = 3400 \text{ Mc/s.}$$

At a carrier frequency of 4000 Mc/s, values varying between 1500 and 2000 Mc/s have been found in practice. It should be remembered in this respect that constructional limitations may give rise to a somewhat larger capacitance than  $C_{ag}'$ .

These remarks apply only to weak signals. At higher power outputs the amplification falls, and in a given amplifier it does so, as may be expected, with the bandwidth remaining almost constant. For example, at a bandwidth of 100 Mc/s the amplification falls 4 dB when the power output is boosted to 1.5 W. In view of the rather low value of the anode load resistance, the power output is mainly limited by the cathode emission.

As an oscillator, the EC 57 at 4000 Mc/s can deliver a power of 3 W with an efficiency of 20%.

The noise factor was measured at various frequencies and was found to be 17 dB at 4000 Mc/s and 16 dB at 3000 Mc/s. At 500 Mc/s it dropped to 5 dB and at 400 Mc/s to 4 dB.

In the development of this microwave triode, important contributions were made by H. J. Lemmens, C. P. Klöpping and C. C. A. M. Moubis. The noise-factor measurements at the lower frequencies were carried out by F. L. H. M. Stumpers.

---

**Summary.** A description is given of a disc-seal triode for microwaves, type EC 57, the dimensions and inter-electrode spacings of which are very much smaller than those of normal triodes. Mention is made of various phenomena which were discovered during the use of microwave triodes (e.g. total-emission damping and noise, and a new type of feedback) and which influenced the design of the EC 57. The EC 57 has planar electrodes connected to chrome-iron discs, between which glass rings are fused (disc-seal construction). A major factor in the design is the use of an L cathode, some advantages of which are a high current density and a metal surface, able to withstand the ion bombardment arising during the degassing of the anode. The disadvantage of the relatively heavy evaporation of barium from this cathode is countered by simple measures, which increase the tube's useful life to about 5000 hours. The 7.5  $\mu$  thick grid wires are wound, under mechanical pre-stressing, on a square frame. Special techniques have been developed for checking the stress so produced and for adjusting with precision the spacing between grid and cathode. The principal tube data are tabulated and the microwave impedances and feedback are discussed. The product of power amplification and bandwidth is approx. 1600 Mc/s and the noise factor at 4000 Mc/s is 17 dB. As an amplifier the tube can deliver 1.8 W at this frequency, as an oscillator 3 W.

## ULTRASONIC MACHINING

### I. TECHNIQUE AND EQUIPMENT

by E. A. NEPPIRAS \*) and R. D. FOSKETT \*).

534.321.9:621.95

*The technique of using high frequency mechanical vibrations for machining brittle materials has assumed considerable importance in recent years. In this technique, a resonant electro-mechanical transducer is used to generate vibrations, at an ultrasonic frequency, which are transmitted to the drilling tool through a mechanical focussing device designed to provide a sufficiently intense vibration at the tool face. The actual cutting agent is an abrasive powder dispersed in a liquid.*

*The growing interest in this ultrasonic machining technique has made it necessary to obtain an accurate assessment of the potentialities of the method. The first part of this article, printed below, deals with the fundamentals of the technique and gives a description of some ultrasonic drilling machines developed at the Mullard Research Laboratories. The second part of the article, to appear shortly, gives an assessment of the technique in terms of cutting speeds, accuracy and surface finish.*

#### Introduction

The first mention of the possibility of using high frequency electromechanical transducers for machining operations is contained in a paper by Wood and Loomis in 1927<sup>1)</sup>. They showed how, by using a piezo-electric crystal vibrator with a focusing device consisting of a tapered glass tube filled with water, holes could be pierced in a glass plate held against the end of the taper. It was not until the war years (1939-45), however, that the technique was put to a useful purpose, when it was applied on a limited scale for cutting and drilling precious stones. Later it was realized that the principle could be extended and applied to other brittle materials, particularly metals and sintered carbides, some of which were found to be readily machined by this method. In 1948 a patent was published<sup>2)</sup> followed by a number of short articles and notices in the American and British press describing some practical results which had been obtained<sup>3-7)</sup>.

The obvious advantage of a reciprocating tool is that the unidirectional motion permits the cavity produced in the workpiece to follow closely the shape of the tool, provided that the tool is restricted to removing particles of material which are small

compared with its own dimensions. In practice, this condition can be fulfilled by restricting the motion of the tool so that the chipping occurs on a microscopic scale, at the same time making the cutting process purely an abrasive one, using the tool itself not as a cutting device but merely to hammer particles of abrasive powder into the work. The tool, aided by the abrasive, impresses its image into the work and it is therefore feasible to make cuts of any required shape by giving the tool the appropriate form. Since reciprocating machines are in this respect more versatile instruments than rotary devices, the field of application of this type of drill is very broad: new types of machining operations are possible, many of which had previously not been attempted. Ultrasonic frequencies are more suitable than sonic frequencies for reciprocating drills not only because operation is silent but also from the point of view of cutting speeds (see below).

#### Fundamentals of the technique

In carrying out a machining operation using the high-frequency reciprocating drill, the tool is pressed into contact with the workpiece using a light pressure superimposed on the alternating motion. An abrasive suspension is fed between the tool and the workpiece. The wearing of the workpiece can be qualitatively explained simply as the result of chipping caused by the abrasive grains being crushed or ground against the work surface by the action of the vibrating tool, the process involving actual contact of the tool with the abrasive particles and the work.

\*) Mullard Research Laboratories, Salfords, Surrey, England.

1) R. W. Wood and A. L. Loomis, *Phil. Mag.* (7) 4, 417-436, 1927.

2) British Patent No. 602 801, 1948, to Industrial Research Corporation, U.S.A.

3) S. G. Kelley, *Materials and Methods* 34, No. 3, Sept. 1951, pp. 92-94.

4) G. H. DeGroat, *American Machinist* 96, Sept. 15, 1952, pp. 141-144.

5) R. G. Woold, *Machinist* 97, 1601, Sept. 26, 1953.

6) E. A. Neppiras, A high-frequency reciprocating drill, *J. sci. Instr.* 30, 72-74, 1953.

7) E. A. Neppiras, *Machining by high-frequency vibration techniques*, *Research (London)* 8, 29-34, 1955.

There are basically two types of operation which can be successfully carried out with reciprocating machines, depending on whether the stock removal is by frictional forces (as in lapping or sizing an existing hole, where the abrasive is rubbed over the surface), or by hammer blows (as in direct piercing or slicing operations). Since in both cases the removal of material is achieved essentially by a chipping action, the technique is limited to comparatively brittle materials and cannot be usefully applied to very soft or merely tough substances. On the other hand, the tool, which is also subject to wear by chipping, is best formed from a tough (not brittle) metal in which the abrasive grains embed themselves without chipping.

The liquid medium holding the abrasive in suspension plays a three-fold role. It acts as a coolant for the tool and workpiece, which would otherwise rapidly become very hot; by capillary action, it allows abrasive to flow to the work area and the worn material to escape; and it achieves a good acoustic bond between the tool and abrasive, allowing an efficient transfer of energy.

Drilling is accompanied by violent cavitation of the liquid between the tool and the work. The audible hiss of cavitation can generally be distinguished above the noise of the actual grinding. The cavitation occurs in the form of streamers of bubbles originating from points on the tool and work. Experiments have shown that the general turbulence produced in the liquid by the motion of these bubble streamers probably helps considerably in stirring up the abrasive mixture under the tool. In this way, by ensuring that broken abrasive is replaced by fresh material and at the same time removing abraded material from the work area, the cavitation action results in an increase in cutting speed. In fact, to a large extent, cutting rates are found to correlate with the observed cavitation intensity.

The cavitation streamers actually consist of multitudes of bubbles in a very violent state of agitation. Bubbles of this sort have an erosive action on solid surfaces and this effect almost certainly accounts for some of the tool wear obtained. The work also often shows wear markings, some of which appear to follow the pattern of the cavitation streamers. They appear to be channels cut into the material merely by the motion of cavitation streamers, carrying abrasive grains with them. The streamers follow fixed paths and the abrasive particles tend to cut comparatively deep furrows in these places. The wear produced in this way is not important from the point of view of stock removal, but it does, of course, affect the quality of the surface finish

obtained and when this is an important consideration, these markings must either be removed after formation or prevented from occurring. This can be done by ensuring that the drilling operations are carried out so rapidly that a stable streamer pattern is not allowed to form.

In the majority of the applications of this technique the three considerations of most practical importance are the cutting speeds, surface finish and machining accuracies obtainable. The most important of these, the cutting speed, is to a large extent dependent on the characteristics of the vibrator itself, e.g. vibration amplitude and frequency, and on the static load between tool and work. Before an efficient drilling instrument can be designed, we must know how these factors are related.

Experience shows that cutting rates depend very much on the nature of the oscillatory motion of the tool. A series of drilling tests in glass showed that under many conditions of operation cutting rates are approximately proportional to the square of the oscillatory amplitude. This is shown in *fig. 1*, where the mean penetration rate is plotted against oscillatory amplitude for each of four drills vibrating at different frequencies, the experimental conditions being otherwise identical in all cases. However, experiments have also shown that, for operations which do not involve frictional forces (lapping) but only direct piercing of the material, there is a practical limit to the usable oscillatory amplitudes: little advantage is to be obtained in making the peak-to-peak oscillatory amplitude of the tool very large compared with the grain dimensions of the abrasive<sup>8</sup>). Moreover, whatever abrasive is used, if the impulse per blow is increased, a point is reached where, because of spraying of the abrasive, it becomes impossible to retain the abrasive paste under the tool at sufficient concentration to allow an efficient cutting action to be obtained. Although these restrictions apply only to direct piercing operations, in order to produce a versatile machine tool capable of both piercing and lapping operations a definite limit must be placed on the oscillatory motion of the tool.

Cutting rates increase somewhat less than proportionately with operating frequency, at ultrasonic frequencies, for constant oscillatory amplitude

<sup>8</sup>) A recent paper by D. Goetze (*J. Acoust. Soc. Amer.* 28, 1033-1037, 1956) on cutting speeds in *tool steel*, shows an approximately linear increase of cutting rates with amplitude. These results, however, refer to peak-to-peak amplitudes up to 0.004", which are about twice the dimensions of the abrasives used. See also Part II of this article, to appear in the following issue of this Review.

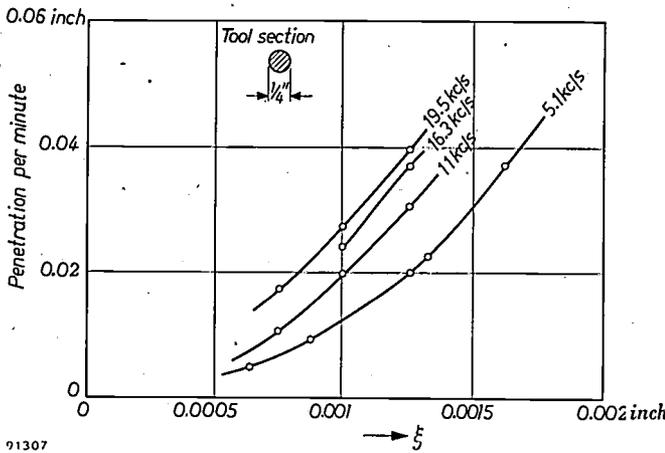


Fig. 1. Cutting rate as a function of oscillatory amplitude  $\xi$  (peak-to-peak) for shallow cuts in glass, at four operating frequencies and for constant static load.

and constant static load. In effect this means that the velocity of the tool is of less importance in determining cutting rates than the oscillatory amplitude. The curve shown in fig. 2 is typical of measured results.

Cutting rates are strongly dependent on the static load imposed between the tool and workpiece and on the tool-face area<sup>9)</sup>, and also on the size and nature of the abrasive particles.

These questions will be discussed in greater detail in Part II of this article. At present, we are interested in the light they throw on the requirements to be fulfilled by the drill vibrator. This consists essentially of two parts: a transducer and a velocity transformer.

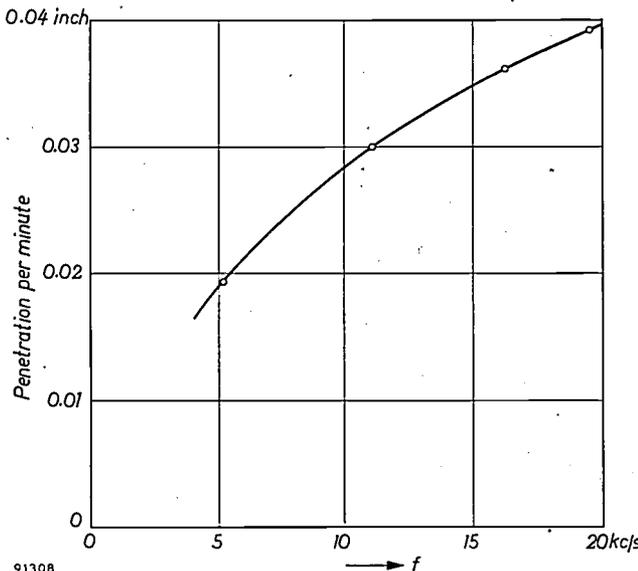


Fig. 2. Cutting rate as a function of operating frequency  $f$ , at ultrasonic frequencies, for a peak-to-peak amplitude of 0.00125" and constant static load.

<sup>9)</sup> See A. Nomoto, J. Acoust. Soc. Amer. 26, 1081-1082, 1954 and also Part II of the present article.

The transducer

The indications are that resonant piezomagnetic<sup>10)</sup> transducers (magnetostrictors) are likely to prove the most valuable of the available types for the present application<sup>11)</sup>. Purely from the point of view of power efficiency a low frequency is an advantage, but there are also obvious advantages in keeping the frequency in the ultrasonic range, from the point of view of inaudibility and to keep the vibrator within reasonable limits of size. It is well known that piezomagnetic resonators are more easily constructed as efficient transducers at low ultrasonic frequencies than any other type of electromechanical transducer. Also, in an application of this sort, the transducers must be capable of withstanding long periods of use under ordinary workshop conditions: piezomagnetic metallic materials are robust and not easily damaged either by rough handling or by the rather high temperatures to which they may be subjected in operation. Under specific pre-stressed conditions, certain recently-developed low-porosity piezomagnetic ceramics<sup>12)</sup> (e.g. nickel-cobalt ferrites) and piezoelectric ceramics<sup>13)</sup> (e.g. lead titanate-zirconates), though somewhat less robust, would also be suitable. Other types of electromechanical transducer such as piezoelectric crystals are more fragile and temperature-sensitive.

As the transducer, only longitudinally vibrating piezomagnetic resonators need be considered here. The most economical form is a consolidated laminar stack driven into resonance by passing a current, at the appropriate frequency, through a coil wound round it, the transducer being biased magnetically by the application of a suitable polarizing field: the optimum biasing point is chosen to give maximum magnetomechanical coupling in the transducer material and corresponds to a flux density about 2/3rds of saturation in most piezomagnetic materials. To achieve this flux density and also to provide a path for the alternating flux, it is generally

<sup>10)</sup> The term *piezomagnetism* is used to represent all reversible, nearly linear magnetomechanical phenomena in polarized ferromagnetic materials, to distinguish them from the irreversible, roughly quadratic effects in unpolarized ferromagnetics. The term *magnetostriction* is used quite generally to include both effects. See for example G. Bradfield, *Acustica* 4, 171-181, 1954, and the second paper referred to in note<sup>12)</sup>.

<sup>11)</sup> For further information on ultrasonic transducers and on ultrasonic techniques generally, see for example, T. F. Hueter and R. H. Bolt, *Sonics*, Chapman and Hall, London 1955.

<sup>12)</sup> U. Enz, *Die Erzeugung von Ultraschall mit Ferriten*, *Tech. Mitt. P.T.T.* 33, 209-212, 1955. C. M. van der Burgt, *Ferroxcube materials for piezomagnetic vibrators*, *Philips tech. Rev.* 18, 285-298, 1956/57.

<sup>13)</sup> W. P. Mason, *J. Acoust. Soc. Amer.* 28, 1207-1218, 1956.

convenient to include a flux return path in the form of a laminated yoke of high permeability, low-loss material. A transducer of this type is shown in the sketch of *fig. 3*.

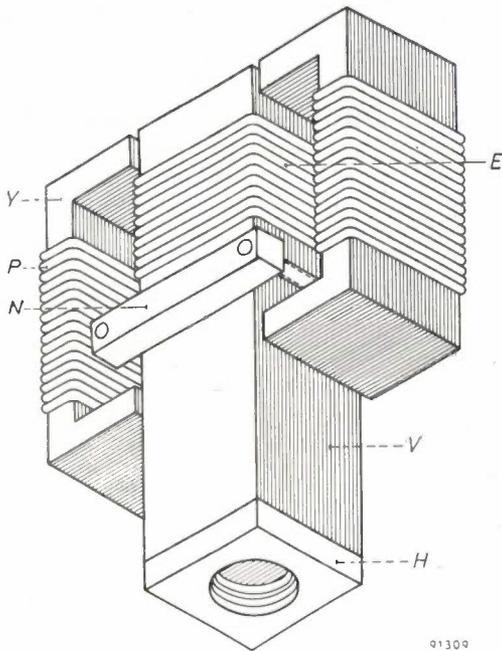


Fig. 3. Essentials of the vibration transducer as used for low-power ultrasonic drills. *Y* laminated yoke, *P* polarizing windings, *N* nodal clamp, *E* energizing coil, *V* laminated vibrating member, *H* tapped steel stub-holder. The vibrating member *V* is  $\lambda/2$  in length.

The additional lamination structure complicates the mechanical design of bar-type structures and becomes inconvenient if the transducer is operated at a power level such that water-cooling is necessary. In such cases, a simpler but less efficient alternative is the window type of construction which is also widely used in ultrasonic work. These transducers consist essentially of two laminated bars joined at the ends. The stack is wound toroidally with a single coil which carries both direct polarization and alternating drive currents. A typical window-type transducer used in ultrasonic drills is shown in the photograph of *fig. 4* (see also *fig. 9*). These transducers are not capable of generating such high vibration intensities as are the bar-type units under comparable conditions of drive<sup>14</sup>). Nevertheless, because of their simple construction and ease of cooling they are generally preferred to bar types in the larger models of ultrasonic machine tools.

<sup>14</sup>) H. H. Rust and E. Bailitis, Kritische Betrachtungen über die linearmagnetostruktive Ultraschall-Erzeugung mittels tonpilzartiger Schwinggebilde, *Akust. Beihefte*, 1952, No. 2, pp. 89-90.

The practical limit, referred to earlier, to the usable oscillatory amplitudes for machining purposes is too large to be attained at the vibrating face of the free transducer. In the transducer, the limit is set by saturation effects and by the danger of fatiguing the piezomagnetic material under the large alternating stresses set up in it. It is, however, possible to increase the amplitude at the tool by directing the vibrational energy on to it through a mechanical focusing device (velocity transformer). In this way maximum cutting rates can be achieved.

### The velocity transformer

The transducer-transformer-tool system which forms the basis of the ultrasonic drill is essentially a resonant mechanical transmission line, two or more half-wavelengths long. The velocity transformer itself generally consists of a tapered metal stub, of suitable high fatigue strength and high mechanical *Q*, which is rigidly bonded to the transducer face. The transformer stubs used in ultrasonic machine tools are generally designed to be resonant at the transducer frequency, as this not only simplifies the design calculations but also allows interchangeable stubs to be used without changing the distribution of nodal planes in the system or its resonant frequency. The importance of this arises from the fact that for the purpose of support it is necessary to provide a rigid clamp, and to avoid excessive damping this should be located at a displacement node; clearly the latter should remain fixed, independent of the stub used and its loading.

The resonant length of tapered metal stubs differs from that for parallel-sided members and depends on the form of the taper and the trans-

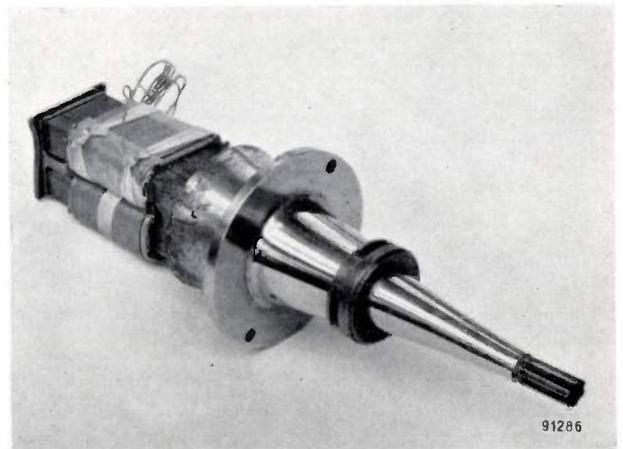


Fig. 4. Transducer-transformer-tool system used on the high-power ultrasonic drill (input power 2 kW, frequency  $\sim 20$  kc/s). The slotted ring clamps the final  $\lambda/2$  stub to the upper mounting stub.

formation ratio required. Calculations are simplified if the stub section is made to change exponentially. In practice therefore this is often done and the design of stubs of this form has been worked out in some detail (see below and <sup>6</sup>). The increase in vibration amplitude  $\xi$  is always exactly equal to the root of the inverse ratio of the end areas for any taper, provided that the maximum lateral dimension of the stub is small in relation to the wavelength. In other words, in any continuous resonant loss-free stub of any number of  $\lambda/2$  sections, the product area  $\times \xi^2$  is constant at all anti-nodal planes and the transformer provides an increase in vibration intensity only at the cost of a reduction in the effective drive face area.

The equation giving the motion of a tuned stub of varying section can be obtained by analogy with the acoustic transformer (or acoustic horn) of which the general equation relating the velocity potential  $\varphi$  with distance  $x$  along the horn and time  $t$  is

$$\frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = \frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial \varphi}{\partial x} \cdot \frac{d}{dx} \ln A, \dots (1)$$

where  $c$  is the velocity of sound in the medium and  $A$  the cross-section at  $x$ . Assuming the vibrations to be simply harmonic, we can write  $\partial^2 \varphi / \partial t^2 = -\omega^2 \varphi$ , where  $\omega$  is the angular frequency, and by substitution in equation (1) we eliminate the time variable and obtain

$$\frac{d^2 \varphi}{dx^2} + \frac{d \varphi}{dx} \cdot \frac{d}{dx} \ln A + \frac{\omega^2}{c^2} \varphi = 0.$$

This equation, which holds for any law of taper, is true only if the largest diameter of the horn is small compared with the wavelength of the vibrations in it and if the "flare" of the taper is not greater than a critical value. For the particular case of the exponential taper, the solution is simple. Here, if  $A = A_0 \exp(-\gamma x)$  where  $A_0$  is the area of the stub at the wide end and the exponent  $\gamma$  defines the flare of the taper,

$$\frac{d^2 \varphi}{dx^2} - \frac{\gamma d \varphi}{dx} + \frac{\omega^2}{c^2} \varphi = 0,$$

or, since the particle velocity  $v = -d\varphi/dx$ ,

$$\frac{d^2 v}{dx^2} - \gamma \frac{dv}{dx} + \frac{\omega^2}{c^2} v = 0,$$

of which the general solution is

$$v = (K_1 \cos \omega x/c' + K_2 \sin \omega x/c') \exp(\gamma x/2), \dots (2)$$

where

$$c' = c/(1 - \gamma^2 c^2/4\omega^2)^{1/2}. \dots (3)$$

Thus the effect of the taper is to increase the effective velocity of sound waves in the material by the factor  $1/(1 - \gamma^2 c^2/4\omega^2)^{1/2}$ .

It can be seen that the velocity is real only when  $\gamma$  is less than  $2\omega/c$ ; this is the critical value referred to above, for the exponential case. Transmission ceases when the flare becomes greater than this value; the horn is an effective filter and this condition determines the cut-off frequency. A consequence of this increase in the effective velocity of sound is that the length corresponding to the fundamental resonance of a tapered stub of this form is greater than that of a uniform rod and is given by  $l = c'\pi/\omega = \lambda'/2$ .

The constants  $K_1$  and  $K_2$  in equation (2) are determined from the appropriate boundary conditions. Considering the fundamental resonance,  $v = v_0$  at  $x = 0$ , where  $v_0$  is the velocity of the transducer face, and the strain is zero at the free end where  $x = l = c'\pi/\omega$ . These conditions give

$$K_1 = v_0 \quad \text{and} \quad K_2 = -\gamma c' v_0 / 2\omega,$$

so that

$$v = v_0 [\cos \omega x/c' - (\gamma c'/2\omega) \sin \omega x/c'] \exp(\gamma x/2). \dots (4)$$

For the half-wavelength stub, at  $x = l$ , the particle velocity  $v_1 = -v_0 \exp(\gamma l/2)$ , so that

$$-v_1/v_0 = (A_0/A_1)^{1/2} = \exp \gamma l/2 = a, \text{ say,}$$

where  $A_1$  is the area of the stub at  $x = l$ . Thus, for a  $\lambda/2$  resonant stub the particle velocities, and therefore amplitudes, are increased at the small end, as compared with the large end, in the ratio of the square root of the inverse area ratio.

The quantity  $a$  mentioned above is a useful design parameter. The effective velocity of the sound waves may be conveniently expressed in terms of the ratio  $a$  of the particle velocities at the small and large ends. Since  $\gamma = (2/l) \ln a$  and  $c' = \omega l/\pi$ , equation (3) may be re-written as

$$c' = c/[1 - (c/c')^2 (\ln a)^2/\pi^2]^{1/2},$$

which rearranged gives

$$\frac{c'}{c} = \sqrt{1 + \left(\frac{\ln a}{\pi}\right)^2} \dots (5)$$

The distance  $x_n$  from the large end to the displacement node may also be expressed in terms of  $a$ . From (4), the particle velocity is zero at the point  $x_n$  given by

$$\cot \frac{\omega x_n}{c'} = \frac{\gamma c'}{2\omega} = \frac{\ln a}{\pi}, \dots (6)$$

and since this is positive,  $x_n < \lambda'/4$ .

The plane of maximum particle velocity is found by differentiating (4) with respect to  $x$  and equating to zero:

$$\frac{\partial v}{\partial x} = v_0 \frac{\gamma}{2} \exp\left(\frac{\gamma x}{2}\right) \left[-\left(\frac{\gamma c'}{2\omega} + \frac{2\omega}{\gamma c'}\right) \sin \frac{\omega x}{c'}\right] \dots (7)$$

This is zero for  $x = 0$  and  $x = \lambda'/2 = l$ . The planes of maximum particle velocity are therefore at the extremities.

The strain  $\epsilon(x)$  in the exponential stub is proportional to (7), since  $\epsilon = \partial \xi / \partial x$  which is proportional to  $\partial v / \partial x$ . Also, for the stress distribution  $\sigma(x)$ , we have

$$\sigma(x) = E\epsilon(x) = E \frac{\partial \xi}{\partial x} = K v_0 \exp\left(\frac{\gamma x}{2}\right) \sin \frac{\omega x}{c'}, \dots (8)$$

where  $K$  is a constant involving  $\gamma$ ,  $c'$ ,  $\omega$  and the Young's modulus  $E$ . Differentiating with respect to  $x$  and equating to zero we get

$$\frac{\partial \sigma(x)}{\partial x} = K v_0 \frac{\omega}{c'} \exp\left(\frac{\gamma x}{2}\right) \left[\frac{\gamma c'}{2\omega} \sin \frac{\omega x}{c'} + \cos \frac{\omega x}{c'}\right] = 0,$$

or, since  $\gamma c'/2\omega = (\ln a)/\pi$ ,

$$\cot \frac{\omega x_\sigma}{c'} = -\frac{\ln a}{\pi}, \dots (9)$$

which gives the position of maximum stress  $x_\sigma$  in the exponential stub. The negative value of  $\cot \omega x_\sigma/c'$  implies that  $x_\sigma > \lambda'/4$ , and comparing (9) with (6) we see that  $x_\sigma$  and  $x_n$  are in fact equidistant from the mid-point of the stub (see fig. 5).

The required drilling tool may be screwed or brazed to the end of the stub, the stub generally being chosen so that the dimensions of its free end

are comparable with the tool size. Consequently, to accommodate a large range of tool sizes it is convenient to provide a range of matching transformers having different transformation ratios and therefore different end diameters. Typical transformer stubs are shown in *fig. 5* with drilling tools attached.

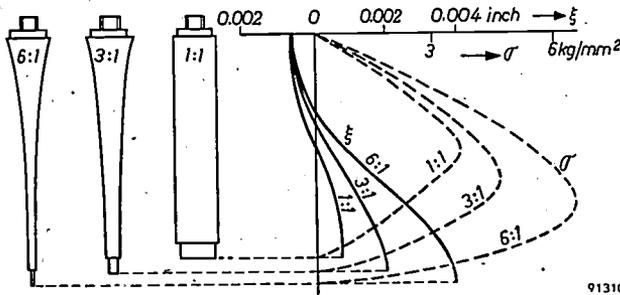


Fig. 5. Half-wavelength velocity transformer stubs with tools attached, showing corresponding amplitude and stress distributions (measured).

A considerable increase in efficiency can be obtained, and the product  $A\xi^2$  at the work-face can be greatly increased, by employing multiple massive coupling stubs and transformers, in which an area discontinuity exists at (displacement) anti-nodal junctions (*fig. 6*). In this way the distribution of vibrational energy in the whole system is re-arranged in such a way that a bigger fraction of the total energy is stored in the less lossy parts, i.e. in the velocity transformer(s). The efficiency of such arrangements depends on the stubs remaining very

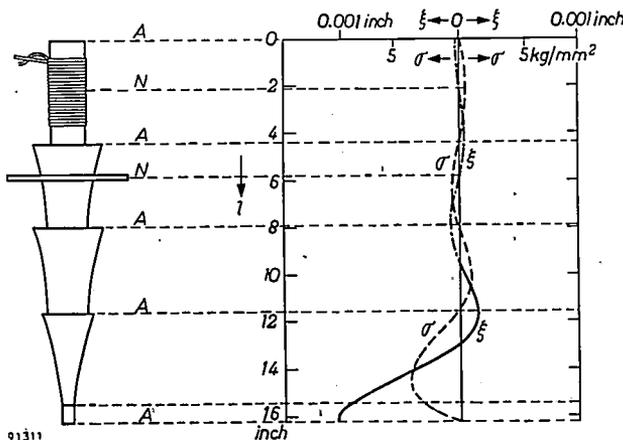


Fig. 6. Transducer-transformer system with step discontinuities at antinodal junctions, showing amplitude and stress distribution (measured). Displacement nodes and antinodes are indicated by N and A respectively.

nearly resonant under all practical loading conditions (so that the stress transmission across the area discontinuity is small). The mechanical  $Q$  of the unloaded system increases and the magnetomechanical coupling coefficient <sup>15)</sup>  $\bar{k}$  is reduced in such a

way that the product  $|\bar{k}|^2 Q$  remains approximately constant. Because of its constancy over a wide range of conditions this product is a very valuable design parameter. When resonant matching stubs are used, the product  $A\xi^2$  referred to the transducer face is proportional to  $(|\bar{k}|^2 Q)^2$  for constant driving conditions if the stubs can be assumed loss-free. If not, then this product decreases with increase in mass of the coupling stubs, being very nearly inversely proportional to the total mechanical damping of the system. In practical cases, therefore, there is of course a limit to the vibration intensity obtainable at the working face under free-running conditions, but this will be very high when the stubs are of low-loss material, corresponding to a high mechanical  $Q$  of the system.

Two other measures may be mentioned by which, under certain conditions, drilling efficiency may be increased above that obtainable with simple continuous stubs.

1) Useful practical mechanical transforming systems can be designed having discontinuities of area in planes other than the displacement antinodes. In particular, a  $\lambda/2$  stub consisting of two  $\lambda/4$  cylindrical sections of different areas (see *fig. 9*) gives an amplitude transformation in the inverse ratio of the end areas (i.e. a greater ratio than the corresponding smoothly-tapered transformer). However, the alternating stress developed in a transformer of this sort is greater than for tapered stubs giving the same drive face amplitude, so that mechanical fatigue troubles may be encountered. The use of these double quarter-wave stubs is therefore restricted to transforming comparatively small end-face amplitudes.

2) As mentioned above, matching stubs must generally be chosen so that the end dimensions are comparable with the tool size. Also a tool loading somewhat below optimum is normally used, in order that small changes in tool face area and loading conditions do not cause too severe mismatching of the amplifier. It is sometimes possible to go some way towards increasing drilling efficiency, however, (without the necessity of increasing the unloaded  $Q$  of the system) by merely increasing the cutting area of the tool in relation to the transformer end area. (In order to take advantage of this, however, the whole electromechanical system has to be designed for one given operation.) In this way the radiation loading due to the liquid medium and the work loading are both increased directly. If the tool area is considerable the radiation loss from the free part of the upper surface may be eliminated by providing acoustic pressure-release material (sponge rubber or a similar material) on this surface.

When using this arrangement it is important to ensure that the dimensions of the tool (its thickness especially) are such that no severe flexural modes can be excited.

<sup>15)</sup> For the definition of the material constant  $k$ , see for example the paper by C. M. van der Burgt mentioned in note <sup>13)</sup>. Apart from this  $k$ , other coupling coefficients may be defined which also take into account the nature and dimensions of the vibrating system, viz.  $k_{y/2}$  (for a  $\lambda/2$  resonator),  $k_{y/2, metal}$  (for a  $\lambda/2$  laminated metal resonator),  $< k_{y/2}$ , owing to eddy current losses) and an overall coupling coefficient  $\bar{k}$  (for a composite system, e.g. resonant transducer + resonant transformer(s) + tool). It is this last,  $\bar{k}$ , with which we are concerned here.

### Ultrasonic drills

As examples of the application of the design principles outlined above, two ultrasonic machining instruments developed and produced by the Mullard Research Laboratories will be described — a small instrument capable of small-scale drilling work which uses a laminated bar-type transducer, and a full-sized machine tool for general workshop use which uses a laminated window-type transducer. The smaller instrument is operated at an input power level sufficiently low that external cooling of the transducer is not needed, while in the larger machine it is necessary to water-cool the transducer. Most of the drilling work described in this paper was carried out with one or other of these machines.

The low-power drill is illustrated in *fig. 7*. This drill is fed from an oscillator supplying 50 W at about 20 kc/s. The transducer (*fig. 3*) consists of a  $\frac{3}{4}$  inch square stack of laminations of nickel which have been given suitable magnetic properties by an appropriate heat treatment. The laminations are bonded together by an insulating cement and fixed to the lower end of the stack is a steel-bush which is tapped to serve as a holder to accommodate a range of velocity transformers. The vibrating

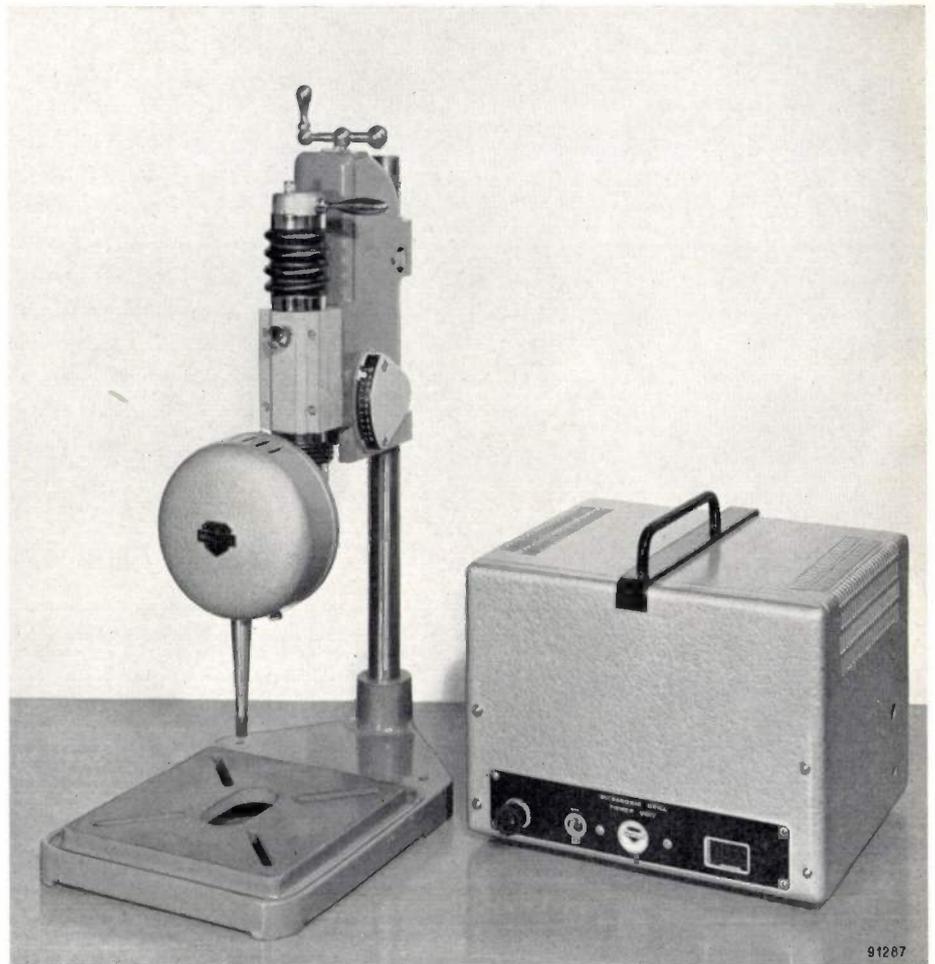


Fig. 7. Low-power ultrasonic drill (input power 50 W, frequency  $\sim 20$  kc/s; Mullard type E 7682).

member is clamped rigidly at its centre (which is a node) and its upper half is surrounded by the energizing coil wound on a former which, to avoid mechanical damping, does not make physical contact with the vibrating laminations. To drive the transducer a conventional oscillator and power amplifier combination is used. A block diagram of the set-up is shown in *fig. 8*. The output from the

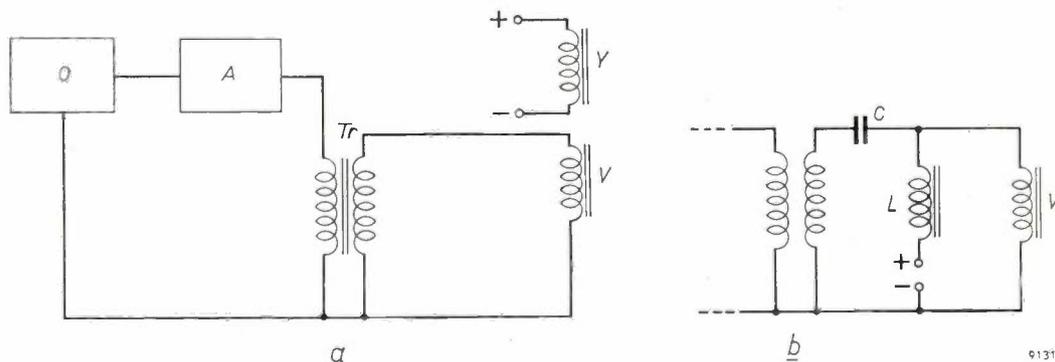


Fig. 8. a) Block schematic diagram of drill vibrator drive system for the lower-power drill. *O* variable frequency oscillator, *A* power amplifier, *Tr* step-down output transformer, + and - terminals for D.C. supply for polarizing field, *V* transducer, *Y* yoke. b) In the high-power drill a blocking capacitor *C* and an L.F. choke *L* are added, whereby drive and D.C. polarizing supply can be fed to a single winding on the transducer.

amplifier is matched, through a step-down transformer designed to handle the chosen frequency efficiently, to the transducer load, which can be varied as necessary merely by adjusting the number of turns on the energizing coil. Polarization for the transducer is provided by completing the magnetic circuit of the transducer with a yoke of high permeability laminations supported sufficiently close to, but not touching, the vibrating member and on which is wound a separate coil carrying direct current from a suitable low-voltage supply as shown in fig. 8a. The velocity transformers, exponentially tapered, which amplify the vibrations obtained at the transducer face, have been shown in fig. 5. The required drilling tip can be screwed or soldered to the end. The complete drill head is mounted in a precision drill stand with arrangements to enable the drill to rest on the work with the necessary light pressure.

The transducer employed in the larger (2 kW) machine is built from nickel laminations of the window type described above. Biasing and drive currents pass through the same winding (see fig. 8b). A square stack ( $V$ ) of these laminations (fig. 9) is brazed to a tapered mounting stub ( $B$ ) and is subsequently consolidated into a solid stack using a resin cement. The complete transducer is supported by a nodal flange at the lower end of the cooling

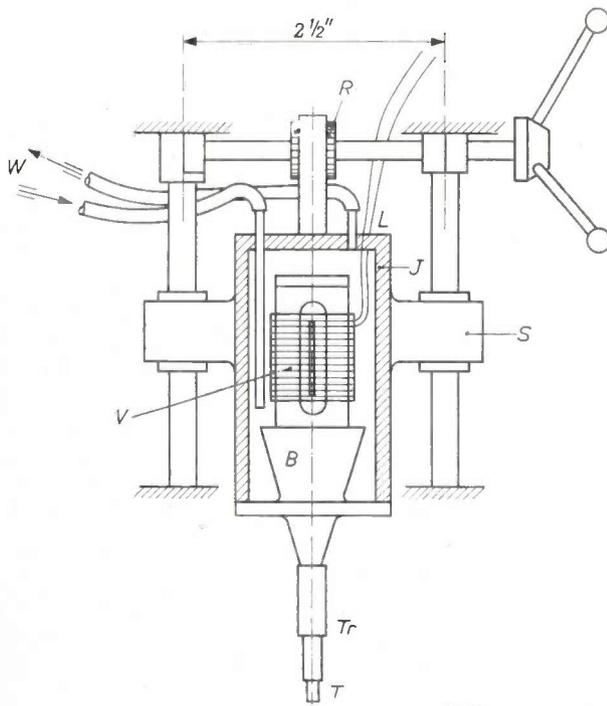


Fig. 9. Essentials of the high-power ultrasonic drill.  $S$  low-friction ball slides,  $J$  water jacket,  $L$  leads to transducer windings,  $R$  rack and pinion mechanism,  $W$  water connections,  $V$  transducer,  $B$   $\lambda/2$  mounting stub,  $Tr$  double  $\lambda/4$  cylindrical velocity transformer,  $T$  tool.

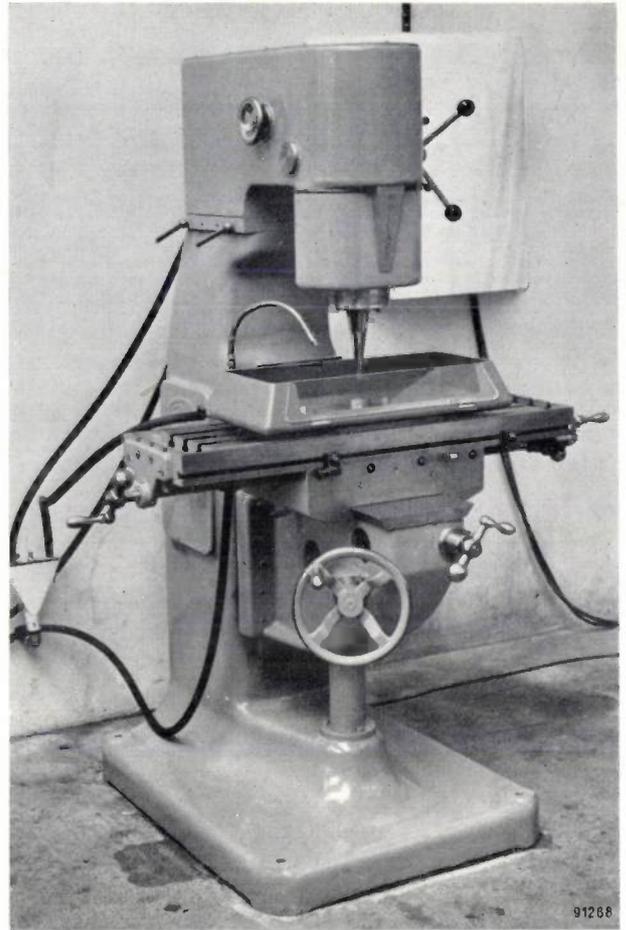


Fig. 10. Mullard high-power drill mounted on a standard type of machine-tool base, with moveable work table.

jacket ( $J$ ) and may be raised and lowered along low-friction ball slides ( $S$ ) by the rack and pinion mechanism ( $R$ ). The weight of the transducer provides too great a static load for many drilling operations, and may be counter-balanced by means of an adjustable lever and weight system (not shown in fig. 9).

The complete drilling head is mounted above a work table closely resembling that of a standard milling machine. This enables the work to be moved accurately along the three rectangular coordinate axes, thus giving the machine some of the qualities of a jig borer.

Tools are carried at the lower end of a second velocity transformer ( $Tr$ ) which screws on to the mounting stub. A photograph of the complete machine is shown in fig. 10. A close-up of a tapered velocity transformer and tool is shown in fig. 11.

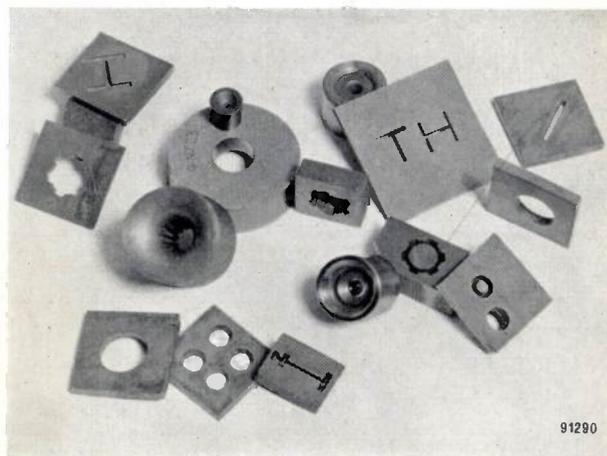
### Applications

The most valuable applications of the vibration machining technique arise from the fact that holes and patterns of intricate shapes can be cut in any brittle material.

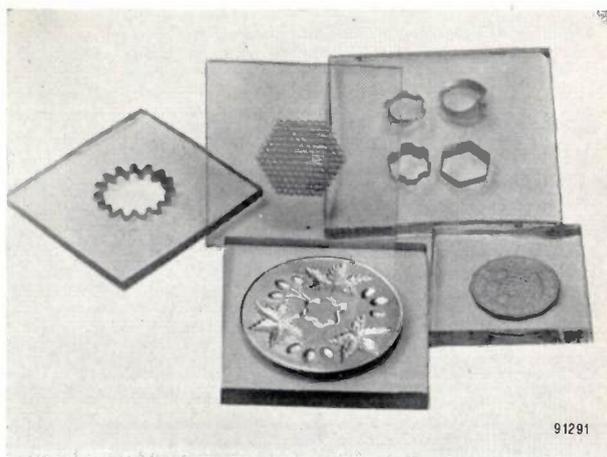
Materials prepared by sintering techniques are often difficult to machine by conventional methods. Normally, sintered parts are pressed to shape before firing, so that the least possible amount of machining has to be undertaken on the sinter. Ultrasonic machine tools enable all manner of intricate shapes to be cut successfully in the fully-sintered state. Applications include the cutting of holes, slots and depressions in ceramics and sintered carbides.

Images and patterns may be simply and rapidly produced by the new method, the tool being merely the reverse impression of the required image, produced in a suitable metal. Since the chipping of the work is on such a small scale, extremely fine detail can be reproduced accurately. For working glass, ceramics and materials of comparable hardness, many impressions can be made from a single tool before tool wear results in any appreciable loss of definition in the image. The technique can of course be used to make dies for medallions or coins, in sintered carbides or die steels. In such materials, tool wear is always considerable so that it is necessary to use both roughing and finishing tools. Fig. 12 shows photographs of cuts in tungsten carbide and glass by this technique.

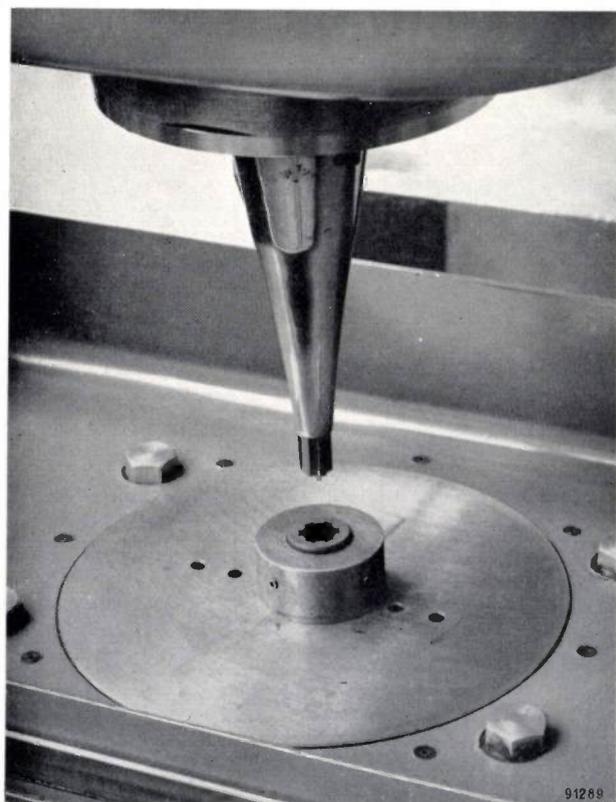
Sapphire, ruby and agate are machined very rapidly; consequently the time required for the manufacture of jewel bearings for watches and



91290



91291



91289

Fig. 11. Close-up photograph of a velocity transformer and tool fitted to the high-power drill.

Fig. 12. Holes and bas-relief patterns of various shapes cut in sintered tungsten carbide (above) and glass (below).

precision instruments can be reduced considerably. In such applications, the major operation consists in cutting out small discs from thin sheet material. As the required dimensions are very small, it is possible and indeed for economic reasons necessary to use a multiple tool capable of blanking out a large number of discs in a single operation. The tool actually used in a typical job of this sort consisted of a cylindrical piece of mild steel rod  $\frac{3}{8}$ " long,  $\frac{5}{8}$ " diameter, having 35 holes each 1.6 mm diameter drilled into it longitudinally. To ensure a parallel-sided cut, and to avoid chipping at the edges as the tool penetrates the work, the sapphire sheets are fixed rigidly to a glass backing plate and the tool is allowed to penetrate not only the work but also some distance into the backing material. In this way discs have been cut to a dimensional accuracy of  $\pm 0.001$  inch on diameter in about one minute from sapphire sheets 1 mm thick.

An important application is in the manufacture of transistors, since the materials used, germanium and silicon, are brittle enough to be easily machined by this method. For slicing bulk crystals, the most economical technique is to use a multiple slicing

tool in the form of a set of parallel blades fixed rigidly in a supporting frame with the required separation. For cutting out the small discs required in transistor manufacture, a multiple blanking technique is used, similar to that described above. This is a valuable production technique, a considerable advance on the use of rotary diamond-impregnated tools.

Considerable difficulties exist in applying conventional techniques to drilling very small holes and depressions in brittle materials. This applies even to round holes. Experiments have shown that extremely small holes, down to at least 0.007 inch diameter, can be cut on the vibration machine provided suitable precautions are taken. To apply the very light static loading required, the drill is preferably made immovable and the load applied by supporting the workpiece on a guided platform, itself supported on a light spring. The load can in this way be set by adjusting the compression of the spring. In cutting small round holes where the circularity is important it is an advantage to rotate the workpiece during the drilling. There is no advantage in rapid rotation and a rate of a few revolutions per second is adequate. The rotation also helps considerably in circulating the abrasive under the tool. As examples of the drilling of small holes to close tolerances we may mention the manufacture of diamond wire-drawing dies and, again, in transistor manufacture.

A useful technique for drilling in places which would normally be inaccessible is to couple the vibrations from the transformer to the tool through a flexible coupling, in the form of a thin wire of a material of suitable high fatigue strength and low loss characteristics. This technique also allows the work to be carried out at considerable distances from the transducer when suitable wire is used. Experience shows that this technique is a very convenient one for drilling small holes rapidly when the operation must be done manually and when positional accuracy is not important. The vibrations are transmitted, with very little attenuation, largely as axial vibrations with only a small shear component, even when the wire is bent so sharply that the radius of curvature is small compared with the wavelength. This suggests an interesting application of this technique: the drilling of *curved* holes — holes with axes bent in any desired way and of any required section. Many actual cuts of this sort have been made in glass.

A field of some promise for ultrasonic techniques is in dental drilling. Investigations have been made in this direction using the above-mentioned wire

transmission lines in conjunction with remote transducers and also using miniaturized transducers.

The vibration machining technique is likely to have its greatest application in the field of die making, since the materials used — hardened steels, sintered carbides and diamonds — are brittle enough to be machined quite rapidly by these methods. Die making by conventional methods is generally a very protracted process involving lengthy periods of lapping, much of which must be done manually using expensive diamond powder and diamond-impregnated tools. Moreover, dies of any complexity must generally be made up in a number of segmental sections and are therefore inherently weak. By using the ultrasonic technique these difficulties are overcome. The most complex shapes can be cut out comparatively rapidly in one piece, using inexpensive boron-carbide abrasive, except where diamonds themselves are to be cut. During the investigations recorded in this article, a great number of holes of all shapes have been cut in tungsten carbide. Extrusion dies and wire-drawing dies require a tapering lead-in and exit portion which can readily be produced on the ultrasonic machine using an appropriate tool of tapering section. However, in large carbide dies where stock removal is considerable, machining time can be saved by sintering the carbide blank roughly in the required form (but undersized on all internal dimensions) and using the vibration technique merely for sizing. If this is done it may be possible to reach the required dimensional tolerances in only one operation.

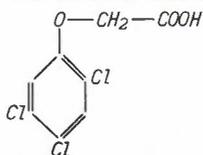
The second part of this article will deal in some detail with cutting speeds, accuracy and surface finish.

**Summary.** This article gives an introduction to the technique of drilling by means of ultrasonic vibrations. The actual cutting action consists of a continuous chipping of the work by abrasive particles in suspension, fed between tool and work. A vibration frequency in the ultrasonic region ( $\sim 20$  kc/s) is used, both for silent operation and in order to get reasonable cutting rates. Piezomagnetic (magnetostrictive) transducers are particularly well suited for generating the vibrations. Two types of laminated metal transducers are described — a bar-type and window-type. These are used in two ultrasonic drills developed by the Mullard Research Laboratories, the former in a small 50 W drill and the latter in a larger water-cooled 2 kW model. The associated velocity transformers are described in some detail: these are resonant stubs, fixed to the transducer and holding the tool, which serve to amplify the vibrations at the tool face. Ultrasonic drilling can be applied only to hard, relatively brittle materials. Among the applications are cutting, the drilling of holes of any cross-section and the production of bas-reliefs, in such materials as glass, tool steel, sintered carbides, ceramics and precious stones. Using a mechanical transmission line of flexible wire it is possible to drill holes in otherwise inaccessible places; even curved holes can be bored. Some of the other applications mentioned are the manufacture of blanking dies and wire-drawing dies and of precious-stone bearings for watches and instruments. The use of ultrasonic machining in transistor manufacture and for dental drilling are also mentioned. A second article will discuss in more detail cutting speeds, accuracy and surface finish.

HERBICIDE MANUFACTURE IN THE PHILIPS-ROXANE FACTORY AT AMSTERDAM



The firm of Philips-Roxane markets a number of insecticides and other chemicals for agriculture and horticulture. One of these is the herbicide 2,4,5 trichloroacetic acid (see inset formula) referred to as 2,4,5-T, used to combat weeds<sup>1)</sup>. The above photograph shows the plant in which this chemical is synthesized. Hexachlorohexane is changed into a dichlorophenol in the autoclave just visible at bottom left of



the photograph. The 2,4,5-T is produced in two further stages: the reaction of the dichlorophenol with monochloroacetic acid and the replacement of a certain H atom in the benzene ring by a chlorine atom. By reaction with suitable alcohols (apparatus to left of steps) the 2,4,5-T is then made into an ester, which is the form in which it is used.

<sup>1)</sup> See Philips tech. Rev. 16, 356, 1954/55; 17, 295, 1955/56.

## SWITCHING TIME OF FERRITES WITH RECTANGULAR HYSTERESIS LOOP

by H. van der HEIDE, H. G. BRUIJNING and H. P. J. WIJN.

621.317.4:538.23:621.318.134

*The use of ferrites with rectangular hysteresis loop as magnetic "memory" elements in electronic computers calls for a means of determining as a function of time the reversal of magnetization caused by current pulses. A method of measuring this process is described in the present article. Some results are given and their possible interpretation is discussed.*

This article on ferrite cores with a rectangular hysteresis loop, as employed in electronic computers, may be regarded as a sequel to a previous article on this subject<sup>1</sup>). That article dealt mainly with the static properties of ferrite cores, whereas we shall now be concerned with their dynamic behaviour. The ferrite cores (fig. 1a) are activated by rapid current pulses which have a more or less square wave-form, i.e. they rise very rapidly (commonly in about 0.2 microsecond) to a certain value, remain constant at that value for some time and then fall rapidly to zero again. When a current

pulse of this kind, shown diagrammatically in fig. 1b, excites a magnetic field in a ferrite, it is found that in most cases the change in magnetization produced by this field cannot follow the increase in the current: this change only begins to become clearly noticeable after the current (the field) has been at its maximum value for some time.

The change of magnetization in the core consists usually in the reversal of the magnetization, for example from negative remanence to the positive magnetization corresponding to the magnetic field applied. (In the following we shall refer to this simply as *reversal*.) The ultimate state of magnetization that is set up, that is, after the passage of the current pulse, is, in the cases in which we are interested, always symmetrical with respect to zero, e.g. from negative remanence to positive remanence.

It is obvious that the duration of the pulse must be longer than the time needed for this reversal, which we shall call the reversal time. If we increase the amplitude of the pulse, the reversal will be more rapid and the reversal time shorter. If we reduce the amplitude the reversal time will be longer. The reversal time does not, however, become continuously longer with decreasing field, for if the magnetic field is too weak, e.g. smaller than the coercive force of the ferrite, the reversal will be only from one non-saturation value to the opposite value and will finally be quite impossible. It is precisely upon this that the practical use of ferrite cores is based (see I). The time for the "non-saturation reversal", which takes place at low field strengths, again appears to be shorter than the longest time needed for almost complete reversal, so that if the reversal time is plotted against the current or the field, a maximum is found in the reversal time. Only that part of the curve at the right of the maximum is of practical importance, for the left part of the curve refers to pulses which do not lead to a full reversal. The maximum reversal time is found approximately at that field strength at which the "rectangularity" or "squareness ratio" (see I) is greatest. For ferrite

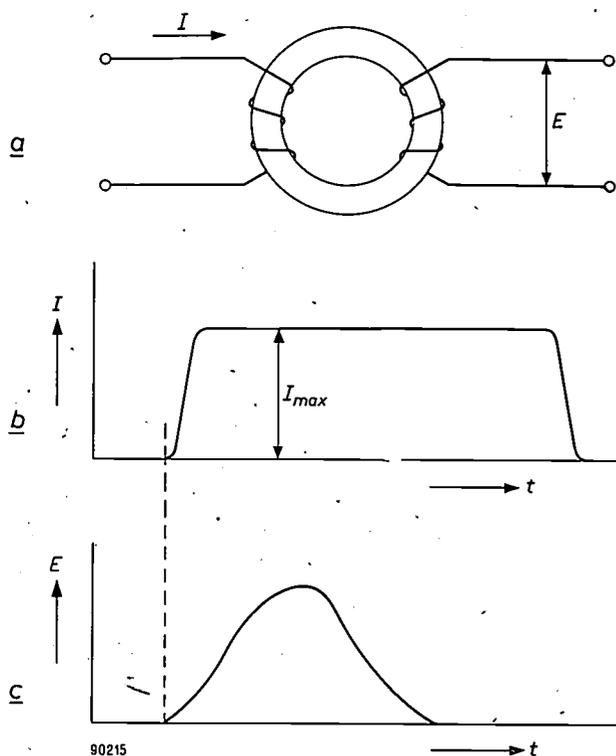


Fig. 1. Ferrite core (a) with two windings. A current pulse (b) in the first winding gives rise to a voltage pulse (c) in the second winding.

<sup>1</sup>) H. P. J. Wijn, E. W. Gorter, C. J. Esveldt and P. Geldermans, Conditions for square hysteresis loops in ferrites, Philips tech. Rev. 16, 49-58, 1954/55; referred to in this article as I.

rings used as memory elements in computers, therefore, a field strength is applied that corresponds roughly to this maximum reversal time.

In the following we shall reserve the term "switching time" for the maximum reversal time.

#### Measurement of the reversal time

The method of measuring the reversal process is in principle as follows. A ferrite ring is given two windings as shown in fig. 1a; a current pulse is passed through the primary winding (fig. 1b) which magnetizes the ring or causes a reversal of the magnetization. As a result of this, a variable voltage is induced in the secondary winding (fig. 1c), the instantaneous value of which is proportional to  $dB/dt$  ( $B$  is the magnetic induction). The area under the curve is thus a measure of the change of induction. The voltage can be viewed directly on an oscilloscope. The measurement naturally requires this process to be repeated at certain regular intervals, and the current pulses employed must be alternately positive and negative. Two primary windings can also be used, wound in such a way that pulses fed to them alternately magnetize the ring first in one direction and then in the other.

The current pulses are commonly generated by valve circuits. Only small ferrite rings can then be activated with adequate rapidity, because the power available is usually rather low. The amount of work needed for a single reversal is proportional to

$$VH\Delta B,$$

where  $\Delta B$  is the change of induction,  $H$  the magnetic field strength and  $V$  the total volume of the ferrite core.

In the investigation described here, measurements were carried out not on the small rings used in memory circuits but on fairly large rings (e.g. 28 mm outside diameter, 20 mm inside diameter and 4 mm thick). Some of these, moreover, had a fairly high coercive force (e.g.  $\mu_0 H_c = 10 \times 10^{-4}$  Wb/m<sup>2</sup> [ $H_c = 10$  Oe]). Since  $\Delta B$  lies almost invariably between 0.2 and 0.8 Wb/m<sup>2</sup> [2000-8000 gauss], the product  $VH\Delta B$  is too large to permit the successful use of normal amplifier valves. For this reason we employed RC discharges via hydrogen thyratrons (type PL 522) for generating the current pulses. In this way we were able to generate pulses of a power more than sufficient for our purposes. These pulses differ from those generated with the aid of vacuum tubes in that although they rise very steeply in the beginning, they have no sharply defined end. As we shall see later, this constituted no objection.

The basic diagram of the circuit used is shown in fig. 2a. Capacitor  $2C$  is charged up to a voltage  $E$  via resistor  $R_t$ . When the thyatron  $T$  strikes, the discharge current  $I$  flows via the thyatron, through the winding of the ferrite ring and the RC circuit.

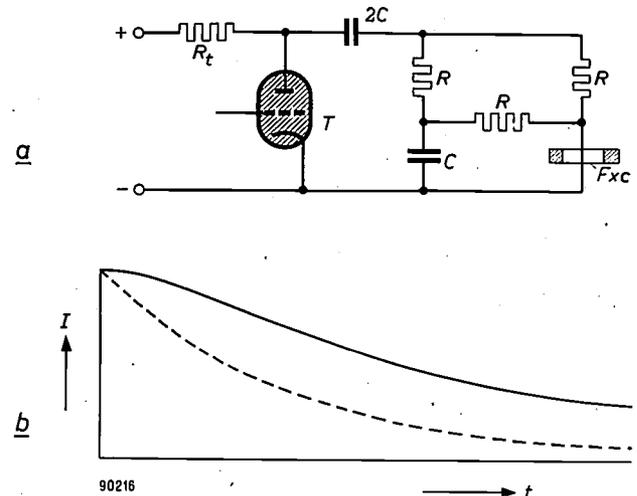


Fig. 2. a) Simplified diagram of the circuit used. When thyatron  $T$  strikes, capacitor  $2C$  discharges and sends a current pulse through the winding on the ferrite core  $F_{xc}$ . b) Dashed: current curve of an ordinary RC discharge according to  $I = (E/R) \exp(-t/2RC)$ . Fully drawn: current pulse supplied by the circuit in (a), according to  $I = (E/R) \left\{ (a+2) \exp(at/RC) + (b+2) \exp(bt/RC) \right\}$ , in which  $a = \frac{1}{2}(-3 + \sqrt{3})$  and  $b = \frac{1}{2}(-3 - \sqrt{3})$ .

The form of the current pulse through the coil is as shown in fig. 2b. If the circuit had only consisted of capacitor  $2C$  and resistor  $R$  on the extreme right, a curve according to

$$I = \frac{E}{R} e^{-t/2RC}$$

would have been obtained, as shown by the dashed line in fig. 2b. Since, however, in the actual circuit capacitor  $C$  is first charged up and subsequently discharged via the ferrite core winding, a flatter current curve is obtained; in fact, with the RC network shown here, the resultant curve is exactly horizontal at the beginning, as shown in the fully-drawn line of fig. 2b. In this case, the current  $I$  falls less than 4% in the time  $\tau = \frac{1}{4}RC$ . For the frequently used values  $C = 6000$  pF and  $R = 6500 \Omega$ , this time comes to approximately 10  $\mu$ s. For almost all the ferrites measured, the reversal times are much shorter than 10  $\mu$ s; in fact, they are usually shorter than 3  $\mu$ s.

As we have seen, field pulses are needed in both directions, and these are obtained by duplicating the arrangement given in fig. 2a. Fig. 3 shows the complete circuit, which functions as follows.

During the half cycle in which the anodes of the diodes are positive, the two capacitors  $2C$  are charged up by the 50 c/s transformer. A special pulse generator then supplies the grid of one of the thyratrons  $T$  with a voltage pulse which causes the

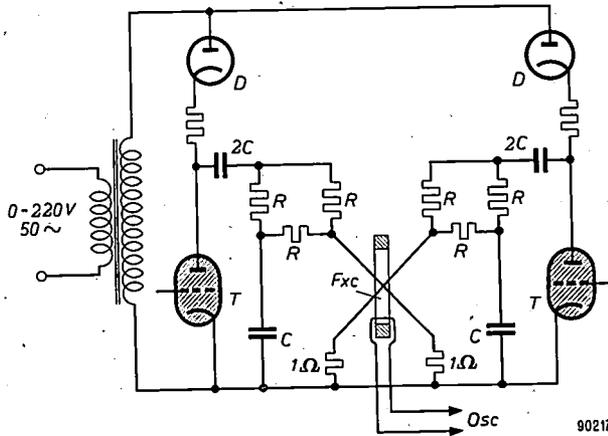


Fig. 3. More detailed circuit diagram.  $D$  diodes, other symbols as in fig. 2. The current pulses are synchronized with the mains frequency, so that no D.C. supply is required. The two halves of the circuit operate alternately and produce successive fields of opposite sign in the ferrite core  $F_{xc}$ .

thyatron to strike, as a result of which a current pulse as shown in fig. 2b flows through one winding of the ferrite ring. One cycle later the same process is repeated with the other thyatron, the result of which is a current pulse in the second winding, which excites an opposing field in the ferrite ring. For the purpose of measuring the voltage pulse there is a third winding, connected directly to an oscilloscope.

The repetition frequency of each pulse train is thus 25 c/s. Now the duration of the voltage pulses may be as short as  $2 \mu\text{s}$ , whilst they are displayed on the oscilloscope at intervals of  $40\,000 \mu\text{s}$ . It would therefore be quite useless to apply the pulse to an oscilloscope with the normal linear time base, generated by a sawtooth voltage. For this reason use is made of a "pulse oscilloscope" (type GM 5660) the time base of which is triggered only during a time interval that is short compared with one cycle of the repetition frequency. The effect of this is the same as if the time base were to extend far beyond the screen. The horizontal movement of the light spot is synchronized with the pulses by means of an extra signal from the external pulse generator mentioned above. From the fact that the electron beam describes a pattern on the screen during only a very small fraction (on an average  $10^{-4}$ ) of the total time, it follows that the pattern cannot be very bright. In an undarkened room the observer must therefore view the patterns under a black hood, in the manner of a photographer.

Having now obtained the oscillogram of the voltage pulse of the ferrite at a given pulse current, we can proceed to measure, for example, the total duration and the maximum amplitude of the pulse. It is often useful, however, to determine the complete variation of the voltage as a function of time. This might be done by photographing the oscillogram, but it is easier to read off a few points of the curve on an illuminated scale, previously calibrated in time and voltage. For adjusting the time scale, a voltage source in the oscilloscope itself was used, which supplies an alternating voltage at a frequency of 1 Mc/s; for the voltage scale a special and very simple square-wave generator was used.

In order to measure the pulse current as well, resistors of  $1 \Omega$  were connected in series with the two primary windings of the ring. The voltages across these resistors were also measured on the oscilloscope. In the RC circuit,  $C$  had the value  $6000 \text{ pF}$  (i.e.  $2C = 12\,000 \text{ pF}$ ) and three values were used for  $R$ , viz.  $12\,000 \Omega$ ,  $6500 \Omega$  or  $960 \Omega$ . These resistors are formed of arrays of carbon resistors, consisting of 8 annular groups of 10 parallel resistors in series. Each array  $R$  is thus cylindrical in form, about 7 cm in diameter and about 30 cm in length. This measure is necessary in the first place to keep self-inductances as low as possible, and also to distribute the dissipated power ( $70 \text{ W}$  at  $15 \text{ kV}$ ) over a large number of carbon resistors. As can be seen from fig. 3, six such arrays are needed for the six resistances  $R$ .

### Results of measurements

Fig. 4 shows three secondary voltage pulses as measured on ferroxcube 6A, for field strengths  $\mu_0 H = 1.3, 1.85$  and  $2.6 \times 10^{-4} \text{ Wb/m}^2$  [ $H = 1.3, 1.85$  and  $2.6 \text{ Oe}$ ]. What is immediately noticeable

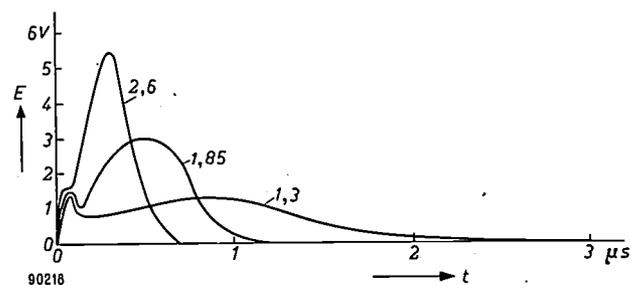


Fig. 4. Voltage pulses measured on ferroxcube 6A. In order of increasing magnitude:  $\mu_0 H = 1.3, 1.85$  and  $2.6 \times 10^{-4} \text{ Wb/m}^2$ .

is that the voltage pulse can show two separate maxima; this points to two different processes in the reversal. For these ferrites with rectangular hysteresis loop the first maximum always represents

a small percentage of the total area under the curve and hence of the change in magnetization  $\Delta B$ . We see that at greater field strength the voltage pulse has a greater amplitude and a shorter duration. The duration  $\tau$  of a voltage pulse is defined <sup>2)</sup> as the time (counted from the beginning of the current pulse) which elapses before the voltage has dropped to 10% of the maximum value; for the maximum value the second peak is considered.

of which is given by the constant  $s$ . The significance of  $H_1$  (often referred to in the literature as  $H_0$ ) is immediately apparent from the figure. In the case of ferroxcube 6A, considered above, the resultant values are  $\mu_0 H_1 = 0.9 \times 10^{-4} \text{ Wb/m}^2$  [ $H_1 = 0.9 \text{ Oe}$ ] and  $s = 0.9 \times 10^{-4} \mu\text{s Wb/m}^2$  [ $0.9 \mu\text{s Oe}$ ].

In cases where  $dB/dt$  after the first sharp peak falls off and shows no further maximum (see fig. 8), it is difficult to decide on clearly defined values of  $\tau$ . It is therefore useful to introduce another definition of the reversal time, in order to be able to define clearly the reversal time for arbitrary forms of voltage pulse, e.g. of ferrites which have no rectangular loop at all. For this purpose we can take twice the time in which the reversal process as regards the change in induction ( $\Delta B$ ) is half completed. The reversal time  $\tau'$  so defined can best be established by determining at what point a time-integrated pulse curve has reached half of its final value (fig. 6a). The integral curve can be determined graphically from a plotted pulse curve, but it is simpler and quicker to observe the electrically integrated curve on an oscilloscope. Fig. 6a shows integral curves found in this way for ferroxcube 6A at a number of pulse field strengths. Fig. 6b gives the reciprocal of the determined values of  $\tau'$  as a function of the pulse field strength  $H$ . It can be seen that, like  $1/\tau$ ,  $1/\tau'$  also shows a minimum. In fig. 6c the total change in induction, which can very easily be found by this method, is plotted against the pulse field strength  $H$ .

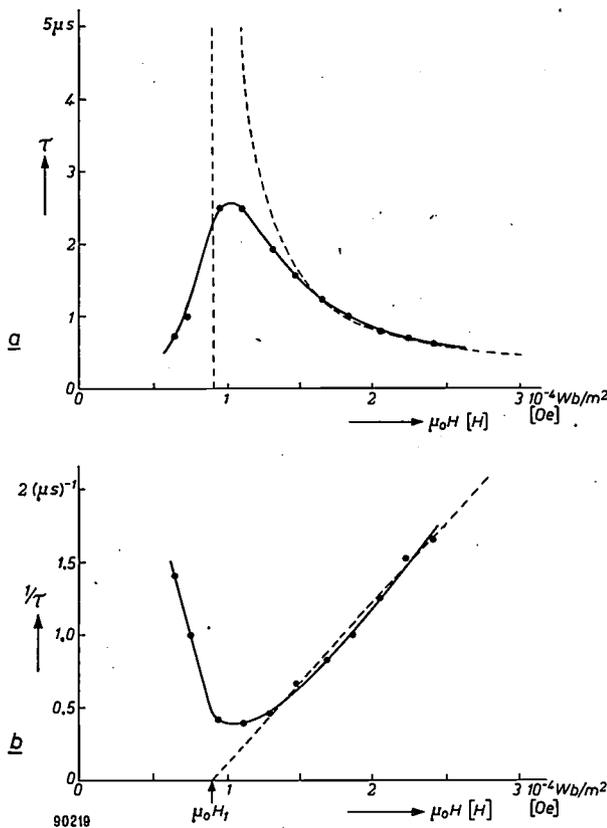


Fig. 5. a) Reversal time  $\tau$  as a function of the pulse field strength  $H$ , for ferroxcube 6A. b) Reciprocal of reversal time,  $1/\tau$ , as a function of  $H$ , for ferroxcube 6A. The slope of the straight line is determined by  $s = \mu_0(H - H_1)\tau$ .

Fig. 5a shows the reversal time as a function of the corresponding field strength  $H$ . We can see the maximum in  $\tau$  mentioned in the introduction. After  $\mu_0 H \approx 1 \times 10^{-4} \text{ Wb/m}^2$  [ $H \approx 1 \text{ Oe}$ ] the curve becomes approximately a rectangular hyperbola (dashed):

$$\mu_0(H - H_1)\tau = \text{const.} = s.$$

If we plot the reciprocal of  $\tau$  as a function of  $\mu_0 H$ , as in fig. 5b, the result is a straight line <sup>3)4)</sup>, the slope

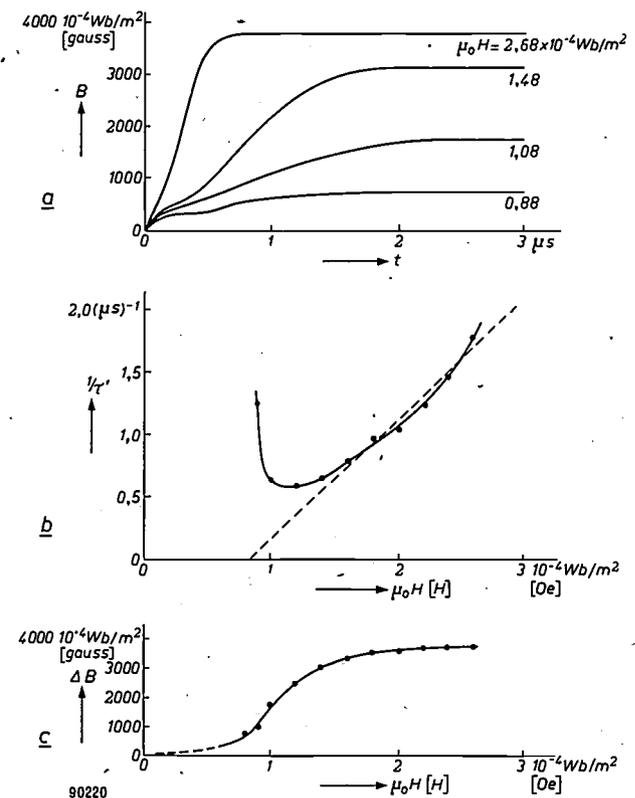


Fig. 6. a) Integral curves  $B = f(t)$  for ferroxcube 6A, for various values of  $\mu_0 H$ . b) Reciprocal of  $\tau'$ , derived from (a), as a function of  $\mu_0 H$ . c) Total change of induction  $\Delta B$  as a function of  $\mu_0 H$ .

<sup>2)</sup> N. Menyuk and J. B. Goodenough, J. appl. Phys. 26, 8-18, 1955.  
<sup>3)</sup> J. K. Galt, Bell Syst. tech. J. 33, 1023-1054, 1954.  
<sup>4)</sup> J. Wylen, Trans. A.I.E.E. 72, Part I, 648-656, 1953.

**Effects of the chemical composition**

In article I some experimental compositions of ferrites with rectangular hysteresis loop were mentioned, together with their various properties. The pulse measurements described in the present article were carried out on these and other ferrites. It is found that the great majority of ferrites with the rectangularity and the coercive force usual for switching elements all have a "reversal constant"  $s = \mu_0(H-H_1)\tau$  of the same order of magnitude. A synopsis of these values is given in Table I.

The most striking deviations from the normal values of  $s$  and  $\tau_{max}$  are shown by  $Co_{0.1}Fe_{0.9}^{II}Fe_2^{III}O_4$  and  $MnFe_2O_4$ . The pulse characteristics of these compounds are given in figs 7 and 8. As a general rule it appears that the switching time is longer the better the rectangularity of the hysteresis loop, i.e. the larger the squareness ratio  $R_{s max}$  (see I). Having regard to this rule,  $Co_{0.1}Fe_{0.9}^{II}Fe_2^{III}O_4$  also fits well into the normal pattern. As a result of the heat treatment in a magnetic field, the squareness ratio in this case is particularly large;

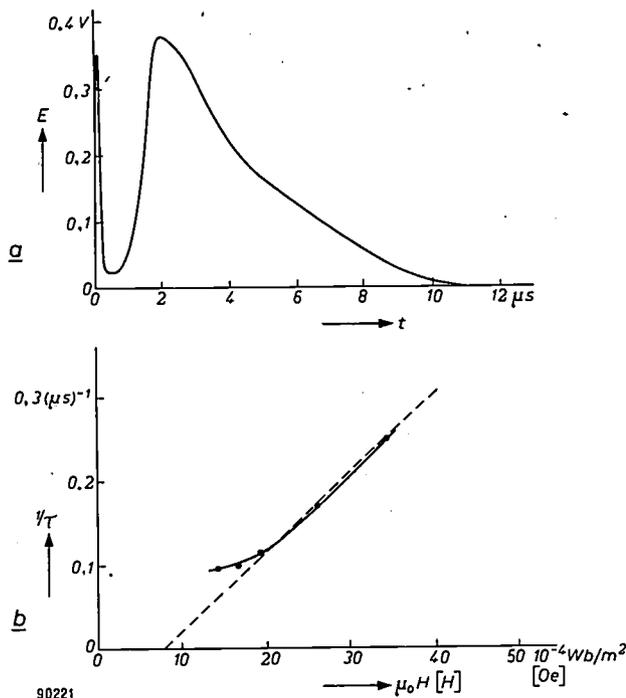


Fig. 7. a) Voltage pulse measured on a core of  $Co_{0.1}Fe_{0.9}^{II}Fe_2^{III}O_4$  at  $\mu_0H = 19.5 \times 10^{-4}$  Wb/m<sup>2</sup>.  
b)  $1/\tau$  as a function of  $\mu_0H$ ; in this case  $s$  is particularly large.

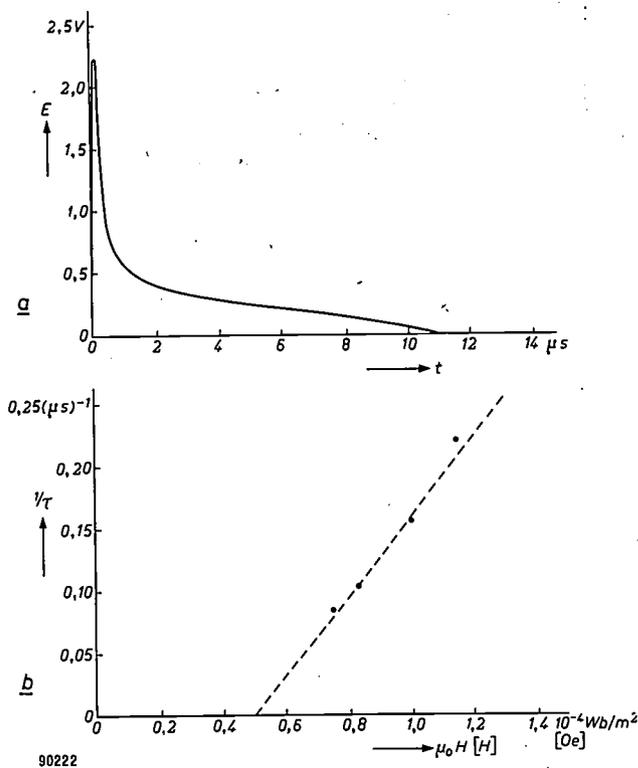


Fig. 8. a) Voltage pulse measured at  $\mu_0H = 0.83 \times 10^{-4}$  Wb/m<sup>2</sup> on a core of  $MnFe_2O_4$ . There is no second peak.  
b)  $1/\tau$  as a function of  $\mu_0H$ .

Table I. Switching times and other properties of various ferrites with a rectangular hysteresis loop:  $\tau_{max}$  longest possible reversal time (switching time); for  $\mu_0H_1$  see fig. 5b;  $s$  slope of  $1/\tau = f(\mu_0H)$ ;  $R_{s max}$  maximum squareness ratio corresponding to induction  $B_m$  obtained at  $\mu_0H_{max}$ ;  $B_r$  induction in remanent state.

Ferrite	$\tau_{max}$ $\mu s$	$\mu_0H_1$ $10^{-4}$ Wb/m <sup>2</sup> [ $H_1$ in oersted]	$s$ $\mu s \times 10^{-4}$ Wb/m <sup>2</sup> [ $\mu s \times$ oersted]	$R_{s max}$	$B_r/B_m$	$\mu_0H_{max}$ $10^{-4}$ Wb/m <sup>2</sup> [ $H_{max}$ in oersted]
Ferroxcube 6A	2.5	0.9	0.9	0.8	0.92	1.5
$Cu_{0.25}Mn_{0.75}Fe_2O_4$	2.2	0.85	0.8			
$(MgO)_{0.5}(MnO_{1+\delta})_{0.875}Fe_2O_3$	1.3	1.4	0.9	0.75	0.88	1.5
$Mg_{0.6}Ni_{0.4}Fe_2O_4$	2.5	3.6	2.2	0.8	0.90	2.2
$Li_{0.47}Ni_{0.06}Fe_{2.47}O_4$	0.6	~ 2.2	~ 1	0.85	0.93	
$Co_{0.1}Fe_{0.9}^{II}Fe_2^{III}O_4$	10	8	105	0.95	0.97	10
Ferroxcube 4A	0.4			~ 0	~ 0.5	0.3
$MnFe_2O_4$	11	0.5	3.2	0.8	0.9	0.5

the same applies to the switching time. The only exception is  $MnFe_2O_4$ . In this case the form of the voltage pulse (fig. 8a) differs considerably from that mostly found. No explanation is yet known for this deviation; it may perhaps have some connection with a systematic inhomogeneity in the ferrite.

**Effect of mechanical stress**

In article I the case was discussed of a glass ring shrunk on to a ring of ferroxcube 4A. As a result of the stress thus exerted on the ferrite, a great improvement was obtained in the rectangularity of the hysteresis loop (see also 5)). The same object was achieved in the present investigation by enclosing a ferrite ring in a rubber ring, and by entirely enclosing the latter in the manner shown in

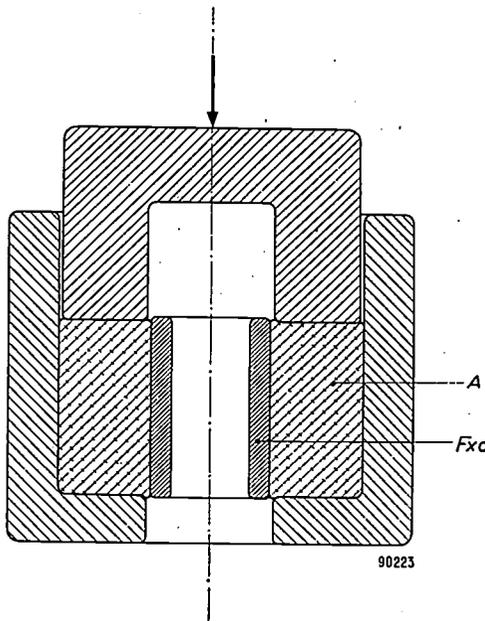


Fig. 9. Schematic diagram of a device for applying a tangential stress to a ferrite core *Fxc* by compressing a rubber ring *A*.

fig. 9 and applying pressure to it. With this method the magnitude of the pressure can only be determined comparatively. In Table II are listed the

Table II. Switching times and other properties of ferroxcube 4A for various values of mechanical stress applied parallel to the direction of the magnetic field.

Stress	$\tau_{max}$ $\mu s$	$\mu_0 H_1$ $10^{-4} \text{ Wb/m}^2$ [ $H_1$ in oersted]	$s$ $\mu s \times 10^{-4} \text{ Wb/m}^2$ [ $\mu s \times$ oersted]	$R_{s \max}$	$B_r/B_m$
None	0.4	—	—	0	0.5
Moderate	1.8	0.6	0.7	0.4	0.7
High	6	1.2	1.8	0.7	0.9
Very high	8	1.6	2.2	0.75	0.95

5) H. J. Williams, R. C. Sherwood, M. Goertz and F. J. Schnettler, Trans. A.I.E.E. 72, Part I, 531-537, 1953.

quantities  $\tau_{max}$ ,  $s$ ,  $R_{s \max}$  and  $B_r/B_m$ ; fig. 10 shows the form of the voltage pulses for a zero pressure and for three increasing values of the pressure. As can be seen, the effect on the squareness ratio

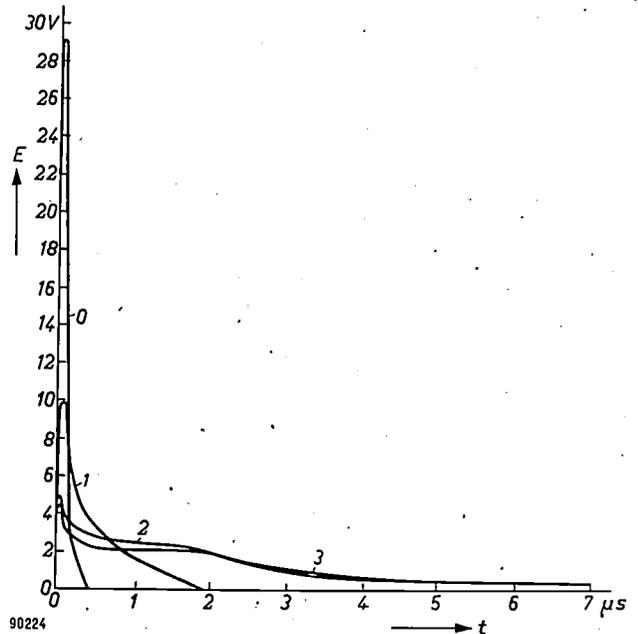


Fig. 10. Voltage pulses measured in ferroxcube 4A at various values of mechanical stress, increasing in the order of curves 0 to 3 and corresponding to the four values given in table II.

$R_{s \max}$  and on the pulse duration  $\tau_{max}$  is surprisingly pronounced. As a result of the pressure,  $\tau_{max}$  is increased by a factor of 20. All this is in good agreement with the above rule that the switching time becomes longer as the hysteresis loop becomes more rectangular.

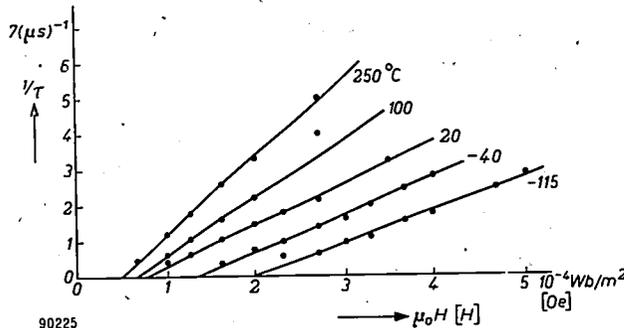
**Effect of the temperature**

It remains to inquire into the temperature-dependence of these switching phenomena. The various properties of ferrites are very differently affected by changes in temperature: for example, the magnetization saturation 6) does not usually change very much in a small temperature range, the initial permeability can change much more, while the electrical conductivity usually undergoes a very pronounced change.

In fig. 11 the reciprocal of  $\tau$  is plotted as a function of  $\mu_0 H$  for  $Cu_{0.25}Mn_{0.75}Fe_2O_4$ , for various temperatures. In fig. 12 the values found for  $s$  and  $\mu_0 H_1$  are plotted against temperature. In the temperature range from  $-100$  to  $+250$  °C the field-strength  $H_1$  changes roughly in the same manner as the coercive force. The reversal constant  $s$  does

6) J. J. Went and E. W. Gorter, The magnetic and electrical properties of ferroxcube materials, Philips tech. Rev. 13, 181-193, 1951/52.

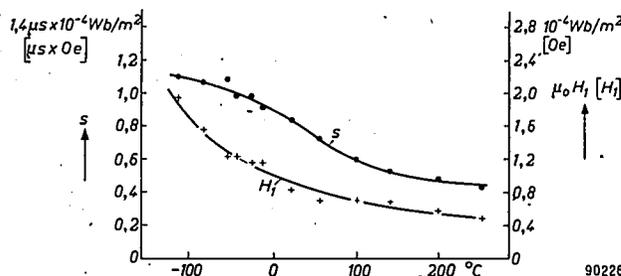
not change much either; it decreases by a factor of 2. This behaviour of the reversal constant is a clear indication that the inertia of the reversal is not due to eddy currents, as it is in metallic ferromagnetic materials. One might at first think that for



90225  
Fig. 11. The reciprocal of the reversal time,  $1/\tau$ , as a function of  $\mu_0 H$ , measured on a ring of  $\text{Cu}_{0.25}\text{Mn}_{0.75}\text{Fe}_2\text{O}_4$  at various temperatures.

this ferrite, eddy currents could be the explanation, in view of the fact that the resistivity  $\rho$  is only  $1500 \Omega\text{cm}$ , which is much lower than in most other ferrites. However, since the resistivity decreases exponentially with increasing temperature, we should expect, if the inertia of reversal were attributable to eddy currents, that  $s$  would rise sharply with temperature, instead of falling slightly. It may be mentioned in this connection that if the effect of eddy currents in metallic ferromagnetic materials is eliminated by using very thin laminations, it is also found that  $s$  varies only very slightly with temperature<sup>7)</sup>.

We shall now inquire into the possible physical explanations of the time effects occurring in connection with reversal.

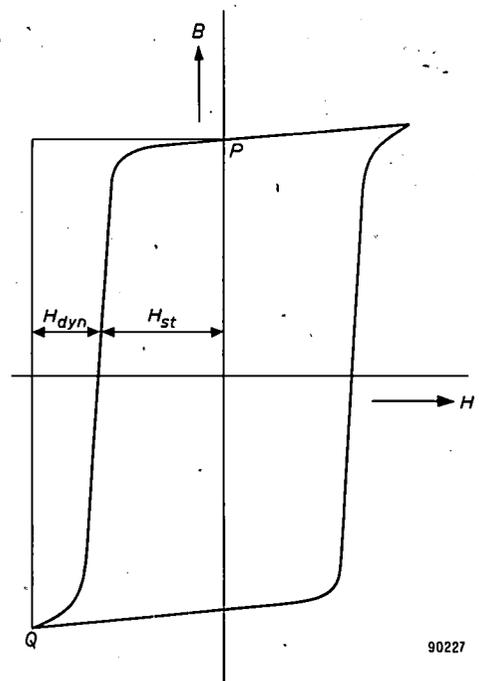


90226  
Fig. 12. Reversal constant  $s$  and  $\mu_0 H_1$  as functions of temperature, measured on a ring of  $\text{Cu}_{0.25}\text{Mn}_{0.75}\text{Fe}_2\text{O}_4$ .

**Interpretation of the observed phenomena**

In attempting to explain these phenomena, we shall first consider the shape of the voltage pulse during the reversal. It is usually found that the secondary voltage shows a sharp peak, followed by

a flatter maximum (figs 4 and 7). This second maximum is sometimes absent, in which case the voltage after the first peak merely falls off gradually and relatively slowly (fig. 8). For a closer analysis we shall start from the normal hysteresis loop in a slowly changing field. We shall divide the field  $H$ , as shown in fig. 13, into a static component ( $H_{st}$ ) and a dynamic component ( $H_{dyn}$ ), and we shall further assume that  $dB/dt$ , and hence the voltage  $E$  of the induced voltage pulse, is at each instant proportional to the instantaneous value of  $H_{dyn}$ .



90227  
Fig. 13. Hysteresis loop indicating the definition of the dynamic field strength  $H_{dyn}$ ; the sum  $H_{dyn} + H_{st}$  is equal to the applied pulse field strength  $H$ .

The reversal takes place from point P to point Q. We can now replot a part of fig. 13 so as to obtain a curve of  $H_{dyn}$  as a function of  $B$  which, according to our assumption, also constitutes a curve of  $dB/dt$  as a function of  $B$  (on a different scale; fig. 14a). We take the reciprocal of  $dB/dt$  and thus obtain a new function (fig. 14b), viz.  $dt/dB$  as a function of  $B$ , which integrated to  $B$  (fig. 14c) gives  $t$  as a function of  $B$ . If the last curve is replotted in the more usual way, with  $t$  as abscissa (fig. 15a), we obtain  $B$  as a function of  $t$ . From this curve we can simply derive  $dB/dt$  as a function of  $t$  (fig. 15b). If our assumption that  $dB/dt$  is proportional to  $H_{dyn}$  is correct, the curve thus derived should correspond to the voltage pulse as obtained by measurement.

<sup>7)</sup> N. Menyuk, J. appl. Phys. 26, 692-697, 1955.

In fact, the agreement of the theoretical pulse form with the measured form is, in most cases, by no means unsatisfactory. The first sharp peak is immediately apparent. (This is also explained by other workers in roughly the same way; see e.g. <sup>8</sup>.) The absence of the rise from zero is due to the fact that, in plotting the curves in figs 14 and 15, we took the ideal case of a rectangular current pulse (zero starting time). Moreover, the theoretical curve

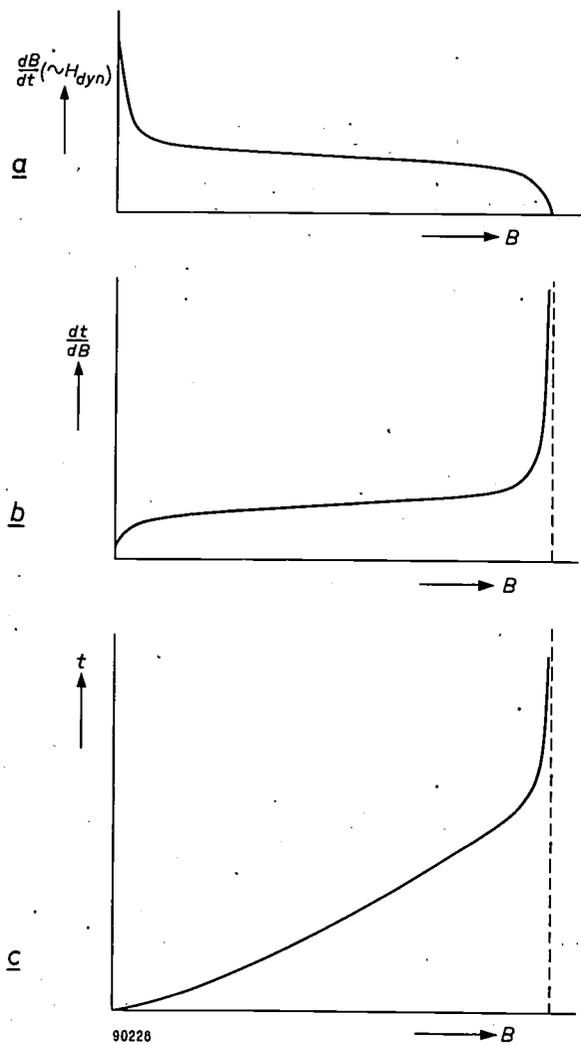


Fig. 14. a, b and c) Graphic derivation of  $t$  as a function of  $B$  from  $H_{dyn}$  as a function of  $B$ , assuming that  $\frac{dB}{dt}$  is directly proportional to  $H_{dyn}$ .

agrees well with the rule that the better rectangularity of the loop, the longer the reversal time (larger region where  $H_{dyn}$ , hence  $\frac{dB}{dt}$ , can be small, fig. 13). However, in no case is there a further rise of  $\frac{dB}{dt}$  after the first peak, whereas in practice this is very frequently observed. We might try to explain this qualitatively by the plausible assumption that the reversal caused by a rapid current pulse takes place according to a mechanism diffe-

<sup>8</sup>) M. Karnaugh, Proc. Inst. Rad. Engrs. 43, 570-584, 1955.

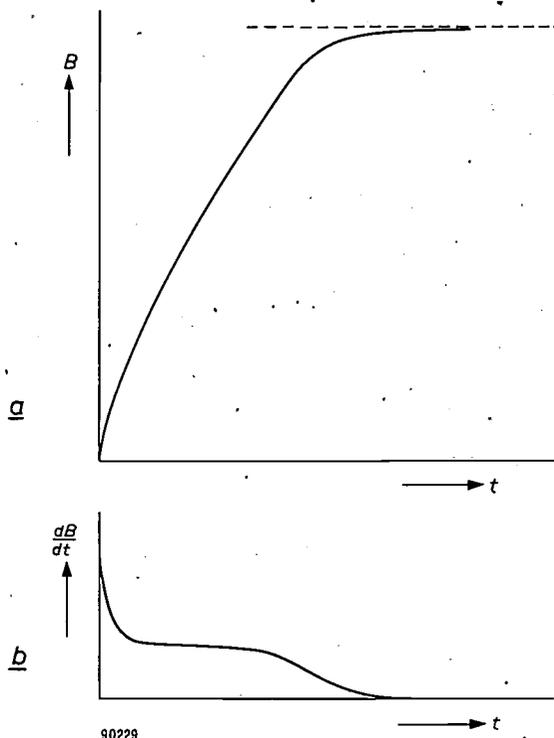


Fig. 15. a) The curve of fig. 14c, replotted with  $t$  as abscissa. b) Form of  $\frac{dB}{dt}$  as a function of  $t$ , derived from (a).

rent from that which applies in a slow reversal (normal hysteresis loop  $H_{st}$  in fig. 13). In the latter case the change of magnetization at each instant is, as a rule, strictly localized, whereas in a rapid reversal the magnetization at a given instant is changing at every point <sup>2</sup>), so that a smaller field than  $H_{st}$  is sufficient in this case, viz.  $H'$ ; see fig. 16.

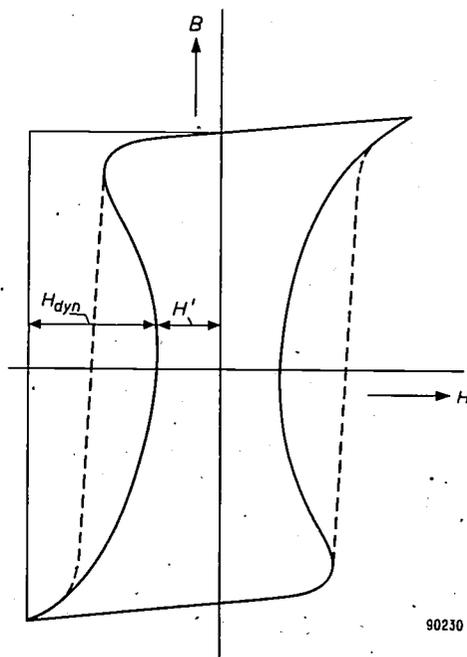
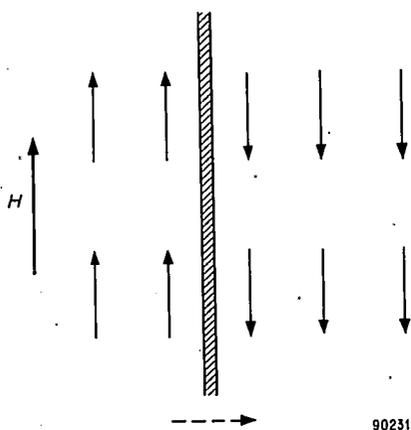


Fig. 16. Illustration of the attempt to explain the second peak in the voltage pulse, assuming that  $H_{dyn} = H - H'$ , and  $H' < H_{st}$ .

### Wall displacements and rotations

We shall now endeavour to go deeper into the effects observed by taking into account the microscopic state of magnetization. We consider the Weiss domains into which a ferromagnetic material is divided and within which the elementary spins<sup>6)</sup> are already in parallel alignment in the non-magnetic state, though being differently orientated in different domains. The direction of magnetization in such domains is usually one of a number of preferred directions with respect to the crystal lattice (see footnote<sup>6)</sup>). These domains are separated by relatively very thin "walls", i.e. transition boundaries, in which the magnetization passes from one direction into another. When, now, a ferromagnetic structure of this kind is magnetized by a magnetic field  $H$ , the domains in which the direction of magnetization does not make too large an angle with the direction of field will grow at the expense of adjacent domains in which this angle is larger; this amounts, therefore, to a displacement of the walls, and it gives rise to a macroscopic magnetization. An example of this is given in *fig. 17* for a frequently considered type of wall known as a  $180^\circ$  wall, in which the spins are parallel with the wall but in opposite directions on either side.



*Fig. 17.* Under the influence of a magnetic field  $H$  the wall of a Weiss domain ( $180^\circ$  wall) shifts in the direction indicated by the dashed arrow.

In domains in which the magnetization is not already parallel or anti-parallel with the direction of field, the magnetization will rotate through a small angle towards  $H$ . Thus, magnetization is also produced as a result of rotation.

Wall displacements constitute by far the larger contribution in a slow reversal of magnetization in a rectangular-loop ferrite. In rapid reversals their contribution is not needed; one can imagine that the wall encounters a certain resistance during

displacement and that therefore it cannot move fast enough to cover a relatively large distance in the required time. Accordingly, with the sudden application of a field  $H$ , the magnetization in the Weiss domain changes only by rotation, and now by rotation through a large angle.

In general, then, there are three possibilities as regards the change in the macroscopic magnetization: it may be caused

- 1) almost entirely by wall displacements,
- 2) almost entirely by rotations or
- 3) by significant contributions of both processes.

Let us return to the first sharp peak, which is almost invariably observed (*fig. 4* and *fig. 7*). This is most probably due to rotations through a small angle. In saying this, we are really making a statement about what follows, because there is no reason why a rapidly beginning rotation should subsequently become so much slower as to give rise to the second peak already mentioned. It seems most probable, therefore, that the further course of the reversal is very largely attributable to wall displacements.

How then is the second peak to be explained? Menyuk and Goodenough<sup>2)</sup> postulate a cylindrical wall, the diameter of which expands at a constant rate after the sudden application of a magnetic field. The rate of wall displacement is thus assumed locally constant; the total  $dB/dt$ , however, is not constant. Since the wall covers an increasing area,  $dB/dt$  increases proportionally to the above-mentioned diameter, and the wall continues to move until it encounters another: if they are of the same type, they then both disappear. New Weiss domains do in fact frequently originate as small "spikes"<sup>9)10)</sup>, which subsequently spread out. The whole process is of course more complicated than described here.

It has further been observed that the second peak becomes higher and narrower at larger pulse field strengths and increasingly tends to merge with the first narrow peak (see e.g. *fig. 4*): There remains then one peak of relatively high voltage and very short reversal time (e.g.  $< 0.2 \mu s$ ), which cannot be measured reliably with the apparatus described. In this case it is quite conceivable that the reversal process takes place mainly via rotations. The ferrite will be strongly magnetized by a strong field pulse of the one polarity; this means that few walls will be left in the remanent state, so that these and subsequently formed walls will have to cover a longer distance when the next pulse of opposite

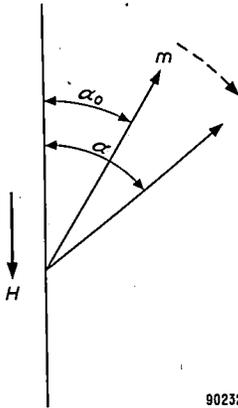
<sup>9)</sup> J. B. Goodenough, *Phys. Rev.* **95**, 917-932, 1954.

<sup>10)</sup> L. F. Bates and D. H. Martin, *Proc. Phys. Soc. A* **66**, 162-166, 1953.

polarity arrives. Also, the field strength may now be sufficient to cause a reversal of magnetization without the intervention of a wall.

This can be verified by a straightforward calculation based on some very simplified assumptions.

We assume that the magnetization vector  $m$  before the field pulse makes an angle  $\alpha_0$  with the field direction as drawn in fig. 18. Upon the sudden application of field  $H$ , the vector



90232

Fig. 18. A magnetic field  $H$  causes the magnetization vector  $m$  of a Weiss domain to rotate in the direction indicated by the dashed arrow.

$m$  rotates in the sense that  $\alpha$  becomes greater. The couple exerted by  $H$  on  $m$  is always

$$M = H m \sin \alpha.$$

We now assume that an opposite couple  $M_t$  arises, which is proportional to the speed of rotation of  $m$  and, of course, to  $m$  itself, since  $m$  is proportional to the volume under consideration. We thus find:

$$M_t = c m da/dt,$$

where the constant  $c$  is a damping constant. At every instant  $M = M_t$ , so that

$$da/dt = \frac{H}{c} \sin \alpha.$$

From this equation it follows that  $dB/dt$  and therefore also the instantaneous voltage  $E$  of the induced pulse, is proportional to

$$\frac{H}{c} \frac{1}{\cosh^2 \left( \frac{H}{c} t + \ln \tan \frac{1}{2} \alpha_0 \right)}$$

An idea of this function is given in fig. 19, in which the appropriate starting angle  $\alpha_0$  is shown at various points. The pulse form obtained is found by postulating the time  $t = 0$  at  $\alpha_0$ . Thus, at  $\alpha = 90^\circ$  there is no longer any peak.

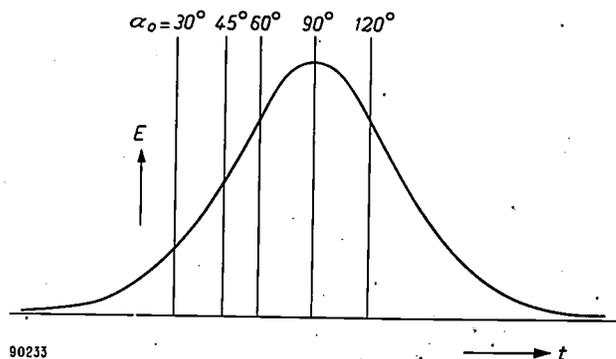
It should be added that a smaller angle  $\alpha_0$  corresponds to a more rectangular hysteresis loop (higher remanence, see also I). Here again, therefore, we find that better rectangularity results in a longer switching time. In reality,  $\alpha_0$  is not, of course, a constant angle, but an average value around which the actual values of the angle are fairly widely spread. Moreover, the crystal anisotropy forces (see I) will also have their effect. If we take into account the fact that the magnetization  $m$  also has a mechanical momentum associated with it (gyromagnetic effect), we nevertheless obtain a very similar result for  $dB/dt$ .

### Effect of mechanical stress

In the foregoing discussion of the effects of mechanical stress on ferroxcube 4A, it was shown that the stressed specimens have improved rectangularity and at the same time a longer switching time: the latter changes from  $0.4 \mu s$  (no stress) to  $8 \mu s$  (maximum stress), i.e. by a factor of 20 (fig. 10). This is in fact more than might be expected from the change in the form of the hysteresis loop. To explain this we must take into account the fact (see I) that in a suitable ferrite under sufficient mechanical stress there are only two preferred directions of magnetization, against 8 or 6 in unstressed ferrite. Accordingly, there are now far fewer possibilities of new Weiss domains being formed; moreover, the number of walls is also reduced owing to the greater anisotropy, so that the walls nevertheless formed must cover greater distances and therefore need more time to do so, the consequence being an extra increase in  $\tau$ .

### Temperature dependence

To conclude, we shall briefly discuss the dependence of the reversal process upon temperature. No matter which mechanism we consider, our calculations will always have to include one damping constant or another. (An example is the above-mentioned constant  $c$ .) If the reversal process can be made to take place exclusively via a certain type of wall, let us say a  $180^\circ$  wall<sup>3</sup>), we shall find for



90233

Fig. 19. The function  $[\cosh^2 (Ht/c + \ln \tan \frac{1}{2} \alpha_0)]^{-1}$ , representing the theoretical voltage curve if the initial value of  $\alpha$  (i.e. at  $t = 0$ ) is  $\alpha_0$ . Various possible starting angles  $\alpha_0$  are shown.

that type of wall a certain damping constant. The question is now, what is it that mainly changes with the temperature, the damping constant itself or the manner in which the reversal process takes place? Changes in the latter with temperature are difficult to imagine in the case of rotations, but

they may well occur in the case of wall displacements. For example, a change in the crystal anisotropy and in the saturation magnetization (see I) will cause a change in the number of the walls. This effect is certainly of influence, although it appears from the article by Galt<sup>3)</sup> that in fact the damping constant also changes with temperature, so that the temperature dependence of the speed of reversal must be partly attributed to this.

**Summary.** A method is described for measuring, as a function of time, the reversal of magnetization caused by the sudden application of a magnetic field to ferrite cores with a rectangular hysteresis loop (used as "memory" elements in electronic computers). Thyratrons are used to produce the powerful current pulses needed for measurements on large ferrite rings. The variation of  $dB/dt$  during the reversal process is measured on an oscilloscope. As a function of time,  $dB/dt$

usually shows a sharp peak, followed by a second, flatter maximum, which is used for defining the "switching time". Some results are set out in graphs and tables. Beyond a certain magnetic field strength the reciprocal of the switching time is a linear function of the field strength. Separate consideration is given to the influence of external stress and temperature. The comparison of different materials as well as the study of the same material under different stresses lead to the rule: the better the squareness ratio, the longer the switching time.

An attempt is made to interpret the phenomena in physical terms, the assumption being made that  $dB/dt$  is proportional to  $(H-H_{st})$ , where  $H_{st}$  is the value of  $H$  as a function of  $B$ , according to the statically measured hysteresis loop. In this way a reasonable explanation is given for the observed variation of  $dB/dt$  as a function of  $t$ . To explain the second peak in  $dB/dt$  an additional assumption is needed. The question is discussed in how far the reversal is effected by rotations of the magnetization vectors in the Weiss domains and wall displacements in these domains; the first sharp maximum in  $dB/dt$  is attributed to rotations, and the second peak to wall displacements. At strong fields (and hence short reversal times) there is often no second peak, in which case the entire reversal is presumably due to rotations.

## ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk \* can be obtained free of charge upon application to the Philips Research Laboratory, Eindhoven, Netherlands.

- 2398:** J. A. van Leeuwen and H. J. Oskam: Simple and sensitive leak detector (Rev. sci. Instr. 27, 328, 1956).

Brief description of a simple and very sensitive leak detector based on the use of outgassed silica gel in conjunction with a Philips Penning gauge. The sensitivity is about  $100\times$  better than that of a palladium tube leak detector.

- 2399:** P. Penning: Annealing of germanium supersaturated with nickel (Phys. Rev. 102, 1414-1415, 1956).

Brief note on experiments to determine the change in density of Ni acceptor levels during the annealing of supersaturated samples of germanium. During the first periods of annealing the density of the Ni acceptor levels decreases much more rapidly than the total density of acceptor levels present in the sample.

- 2400:** J. D. Fast: Moleculaire interpretatie van relaxatieverschijnselen (Chem. Weekblad 52, 445-460; 1956). (Molecular interpretation of relaxation phenomena; in Dutch.)

This article is substantially the introductory address to a symposium held in Leiden in December 1955 on the subject indicated in the title. Relaxation phenomena play an important role in many varied branches of physics and chemistry, where the

mechanical, dielectric or magnetic properties of the material are involved. A section dealing with the phenomenological aspects of relaxation effects, considers such concepts as loss angle, compliance (compressibility, reciprocal elastic modulus, electric and magnetic susceptibility), logarithmic decrement, relaxation and hysteresis. In a subsequent section, various actual relaxation effects are interpreted on an atomic basis. The analogy between mechanical, dielectric and magnetic relaxation is not restricted to the formal mathematical similarity: it is also often manifested in the underlying atomic mechanism. In the final section, the origins of mechanical relaxation and hysteresis in metals are considered; they are caused by movements of atoms, dislocations and domainwalls, heat transport and the shearing of crystal boundaries.

- 2401:** O. Bosgra and H. Hoogendoorn: Enting en immuniteit bij hondeziekte en hard-pad disease van fretten (T. Diergeneeskunde 81, 649-658, 1956). (Inoculation for immunity from distemper and hard pad in ferrets; in Dutch.)

Ferrets vaccinated with 1/50th to 1/300th dog's dose of canine distemper vaccine cultivated on embryonated hens' eggs (N.V. Philips-Roxane), prove to be completely immune against intraperitoneal infection with large doses of virulent

distemper — hard-pad virus. After vaccination with 1/900th dose, one in three ferrets proves not to be immune, the other two are. However, two ferrets vaccinated with 1/2700th dose did prove to become immune. An amount of 1/8100 dog's dose proved to be too small to immunize ferrets. From this, it can be concluded that some animals are hard to immunize with small doses of vaccine. It is therefore advisable that the distemper vaccine contains an ample quantity of active virus. After storage in a refrigerator for 9 months, it appeared that 1/200th dog's dose still gave ferrets a solid immunity. 1/600th dose did not. Vaccine containing mainly inactivated virus gives ferrets insufficient immunity when used in doses of 1/50th dog's dose. Ferrets do not show any disturbing symptoms after vaccinations, not even when vaccinated with 1 dog's dose.

**2402:** O. Bosgra and H. Hoogendoorn: De enting van nertsen tegen infecties met distemper- en hard-pad-virus (T. Diergeneskunde **81**, 659-666, 1956). (Inoculation of mink against infection with distemper and hard-pad virus; in Dutch.)

Experiments are described in which mink were immunized with a living, avirulent canine distemper vaccine, cultivated in embryonated hens' eggs. Varying doses of the vaccine were administered, viz. 2, 1/5, 1/50 and 1/200 times a dog's dose, without causing adverse results. 25 days after vaccination the animals were infected; some with a virulent distemper virus, and some with virulent hard-pad virus.

The vaccinated animals proved to be completely immune. Of the other animals which were unvaccinated and infected for purposes of comparison, those infected with distemper all died showing the typical symptoms of the disease. Of those infected with hard-pad disease, some died and some recovered after having been seriously ill. It is apparent from these results together with the literature also reviewed, that canine distemper virus vaccine cultivated on embryonated hens' eggs, produces an excellent immunity in mink against both distemper and hard-pad disease.

**2403:** H. J. J. van Boort, N. Warmoltz and J. E. Winkelman: Human retina with a discharge flashlamp (Med. and biol. Illustration **6**, 166-170, 1956).

The article describes a quartz discharge flashlamp of maximum efficiency, by means of which a very

effective light source is obtained for colour photography of the retina using the Zeiss-Nordenson camera.

The lamp is mounted end-on to the condenser lens of the apparatus and is operated at 500 volts giving an energy of 750 watt secs. The flash duration for which the light intensity is more than 50 per cent of the peak-intensity (half-peak time) is about 3.5 msec. The maximum brightness is 1.5 million stilbs and the integrated brightness 7400 stilb secs.

Some photographs of the retina taken with the flashlamp and the carbon arc are shown to demonstrate the increase in sharpness that can be obtained with the flashlamp. With smaller energy values the same lamp can be used for photographing the external eye.

**2404\*:** J. M. Stevels and A. Kats: La représentation systématique des imperfections dans les réseaux silicium-oxygène (Verres et Réfractaires **10**, 129-134, 1956).

Translation into French of abstract No. R 290.

**2405:** A. Kats and J. M. Stevels: L'effet des rayonnements X et U.V. sur les verres silicatés, la silice fondue et le quartz cristallin, I (Verres et Réfractaires **10**, 135-150, 1956).

Translation into French of abstract No. R 291.

**2406\*:** K. Teer: Uebertragungssysteme für das Farbfernsehen (Nachr. techn. Z. **9**, 196-198, 1956).

Abbreviated version of abstract No. 2381.

**2407:** P. Westerhof and J. A. Keverling Buisman: Investigations on sterols, VI. The preparation of dihydrotachysterol<sub>2</sub> (Rec. trav. chim. Pays-Bas **75**, 453-462, 1956).

The preparation of dihydrotachysterol<sub>2</sub> to be used as a standard in biological experiments was studied. The best yields were obtained by reduction of tachysterol<sub>2</sub> with excess alkali metal, dissolved in liquid ammonia or suspended in a secondary amine. A laboratory method for the preparation of tachysterol<sub>2</sub> is given.

**2408:** F. A. Kröger: Some optical and electrical measurements on blue fluorescent ZnS-Cl single crystals (Physica **22**, 637-643, 1956).

For hexagonal ZnS-Cl single crystals, measurements of the intensity of luminescence, photo-conductivity, Hall effect, thermo-luminescence and thermo-conductivity have been carried out (the latter two phenomena being the luminescence and the change in conductivity on slowly warming

to room temperature after U.V. irradiation at 90° K). Analysis of the experimental results provides the following data:

a) the electron mobility at room temperature  $\mu_c \approx 120 \text{ cm}^2 \text{ V}^{-1} \text{ sec}^{-1}$ ; the mobility decreases with increasing temperature; b) the depth of traps formed by  $\text{Cl}^-$  is  $\epsilon_{\text{Cl}^-} = 0.24 \pm 0.02 \text{ eV}$ ; c) the frequency of escape from  $\text{Cl}^-$  traps  $s = 7 \times 10^4 \text{ sec}^{-1}$ ; d) the capture cross-section of the  $\text{Cl}^-$  traps is  $\sigma \approx 3 \times 10^{-21} \text{ cm}^2$ ; e) recombination levels ("killers") are situated  $\sim 0.7 \text{ eV}$  below the conduction band.

**2409:** G. H. Jonker: Magnetic compounds with perovskite structure, IV. Conducting and non-conducting compounds (*Physica* **22**, 707-722, 1956).

A further experimental contribution to the knowledge of the magnetic interactions in compounds with perovskite structure is given.  $\text{LaMnO}_3$  is an antiferromagnetic compound, but it is shown that it would be ferromagnetic if the cubic perovskite structure, present at high temperatures, could be preserved at low temperature. Besides the positive magnetic interaction between  $\text{Mn}^{3+}$  and  $\text{Mn}^{4+}$  ions, an interaction of the same sign is found between  $\text{Mn}^{3+}$  and  $\text{Cr}^{3+}$  ions. A survey is given of the results of this and former investigations.

**2410:** G. L. Meijering: Modellversuche über das Entstehen des Stapeldrahtgefüges in Wolfram (Warmfeste und Korrosionsbeständige Sinterwerkstoffe, 2, pp. 305-312, Plansee Seminar, 19-23 June 1955, Reutte, Tirol, edited by F. Benesovsky, 1956). (Model experiments on the origin of non-sag structure in tungsten; in German.)

Certain additions are known to have a marked effect on the metallographic structure of annealed tungsten wires. For example the technically important "non-sag structure" is characterized by long, rather irregular, grain boundaries making on the average a small angle with the wire axis.

Experiments were carried out with two-dimensional models consisting of 30 parallel fibres. Growth obstructions in the fibres and "leaks" in the fibre walls were distributed statistically. Thereupon "crystal growth" resulted in structures quite similar to the experimental ones. This is in accord with the view that remnants of the addition play an essential role in the realization of the non-sag structure in tungsten wires.

**2411:** G. L. Meijering: Erhöhung der Warmfestigkeit durch innere Oxydation von Legierungen (Warmfeste und Korrosionsbeständige Sinterwerkstoffe, 2, pp. 405-410, Plansee Seminar, 19-23 June, 1955, Reutte, Tirol, edited by F. Benesovsky, 1956). (Increase of hot-strength of alloys by internal oxidation; in German.)

Recrystallization, grain-growth and creep are strongly inhibited by internal oxidation if the oxide is formed in a very fine dispersion. Coarse oxide particles have less influence on these high-temperature properties, though more than on room-temperature hardness.

**2412:** B. Verkerk: Autoradiography in study of surface phenomena (*Nucleonics* **14**, July 1956, pp. 60-64).

Notes on the technique of making surface autoradiographs of metal specimens. An example is given in which the effect of chlorine on semiconducting alumina layers is studied.

**2413\*:** H. J. G. Meyer: Ionenschwingungsprobleme bei Uebergängen lokalisierter Elektronen in Halbleitern (*Halbleiterprobleme III*, edited by W. Schottky, pp. 230-260; F. Vieweg und Sohn, Brunswick 1956). (Influence of lattice vibrations on transitions of localised electrons in semi-conductors; in German.)

Abbreviated version of the author's thesis, these abstracts No. 2374\*.

**2414:** A. Claassen and L. Bastings: The photometric determination of copper by extraction as the diethyldithiocarbamate. Interferences and their elimination (*Z. anal. Chem.* **153**, 30-38, 1956).

Two procedures for the photometric determination of copper as the diethyldithiocarbamate complex have been investigated as to their selectivity. In the first procedure — extraction by chloroform from a citrate solution of pH about 8.5, containing EDTA — the following elements interfere:  $\text{Hg}^{\text{II}}$ , Ag, Au, Pt, Pd, Os,  $\text{Sb}^{\text{III}}$ ,  $\text{Te}^{\text{IV}}$ ,  $\text{Tl}^{\text{III}}$  and Bi. Interference by  $\text{Hg}^{\text{II}}$ , Ag, Pd,  $\text{Sb}^{\text{III}}$ ,  $\text{Te}^{\text{IV}}$  and  $\text{Tl}^{\text{III}}$  can be prevented by simple measures. In the second procedure — extraction by a solution of lead diethyldithiocarbamate in chloroform from ammoniacal citrate — only Ag,  $\text{Tl}^{\text{III}}$ , Bi and  $\text{Hg}^{\text{II}}$  interfere. Interference by the first three elements can be prevented easily, leaving as the only serious interference in this method mercury in amounts above 500  $\mu\text{g}$ .

# Philips Technical Review

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

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

## A HIGH-SPEED UNISELECTOR FOR AUTOMATIC TELEPHONE EXCHANGES

by J. M. UNK \*).

621.395.343

---

*The employment of high-speed selectors is of interest in both direct and indirect systems of automatic telephony. In this article a very rapid uniselector is described. A new method of coupling it to a common driving system makes it possible to bring the selector to a stop, against the desired outlet, from a speed of rotation of 300 outlets per second. This speed permits the adoption of novel wiring schemes whereby wear is very considerably reduced.*

---

### Introduction

The selector is an essential element in a telephone system. Basically it is a multi-way switch, the brushes or "wipers" of which are able to travel over a number of contacts in order to establish the connection desired by the subscriber.

Some time ago in this journal a private automatic branch exchange was described which made use of the U45a<sup>1)</sup> high-speed uniselector recently developed by Philips Telecommunicatie Industrie. The employment of this selector has great advantages, not only in private exchanges but also in public ones, both in systems without registers (direct systems) and in those with registers (indirect systems). The selector in question will now be described and in some respects compared with other types, including the two-motion selector. The most interesting feature of the U45a, viz. its great speed, will also be given closer attention.

The construction of the two-motion selector formerly in common use, with its  $10 \times 10$  contacts (outlets), is such that the wipers have only a relatively short distance to travel in order to reach any desired contact; this permits the finding time

to be short and the average speed in steps per second to be relatively low. For example, in order to reach the 96th outlet from the home position, the wipers have 9 vertical and 6 horizontal steps to take, a total of 15 steps. The two-motion selector has the disadvantage, however, that the wiper mechanism is subjected to high accelerations and decelerations, resulting in severe mechanical shocks, heavy wear and high contact noise. A second disadvantage, the seriousness of which will become clear in the course of this article, is that the 100 outlets must invariably be divided into 10 groups of 10, no other groupings being possible.

The use of "one-motion" rotary selectors (uniselectors) having the same number of outlets gives much greater freedom as regards the grouping of the outlets; in a uniselector they can be distributed between various groups in any desired manner. However, in order to obtain finding times equal to or shorter than those of the two-motion selector, the speed of the uniselector wipers must be very much greater, because generally a greater distance has to be travelled. In certain circumstances it is possible to contrive that the average distance to be travelled by the wipers of the high-speed uniselectors is no greater than that for the two-motion selector. We shall come back to this point later.

\*) Philips Telecommunicatie Industrie, Hilversum, Netherlands.

<sup>1)</sup> B. H. Geels, A private automatic branch exchange using high-speed uniselectors, Philips tech. Rev. 18, 19-30, 1956/57.

In order that the U45a<sup>2)</sup> with its higher speed might be provided with a steady and powerful continuous drive, it was decided to run it from a central powering system, in other words, from a motor common to a number of selectors, each selector being driven via a coupling mechanism. The mechanism has only light-weight parts to set into motion and to arrest; in consequence, wear and selector noise are considerably reduced. Motor-driven uniselectors having a speed of 200 contacts per second have previously been developed during the last few decades. The drive of the U45a selector differs radically from that of its predecessors, however. In the U45a the speed of wiper travel has been raised to 300 contacts per second. We shall demonstrate that it is of great importance that the speed should have been raised to this particular figure, in that it permits the adoption of a new method of connecting the outlets whereby, despite the greater speed, a reduction of wear is obtained.

#### Positioning of a group selector in a direct system

In direct telephone systems the selectors of the various stages or steps are positioned directly, one after the other, by the impulses coming from the subscriber's dial. The functions of the various selecting steps, such as final and group selectors and line finders, are briefly

explained by the diagram in *fig. 1* and the accompanying description (see also the articles cited in footnotes<sup>1)</sup> and<sup>2)</sup>).

In direct systems it is necessary that the group selector should have found an idle outlet of the desired group before the following digit is dialled.

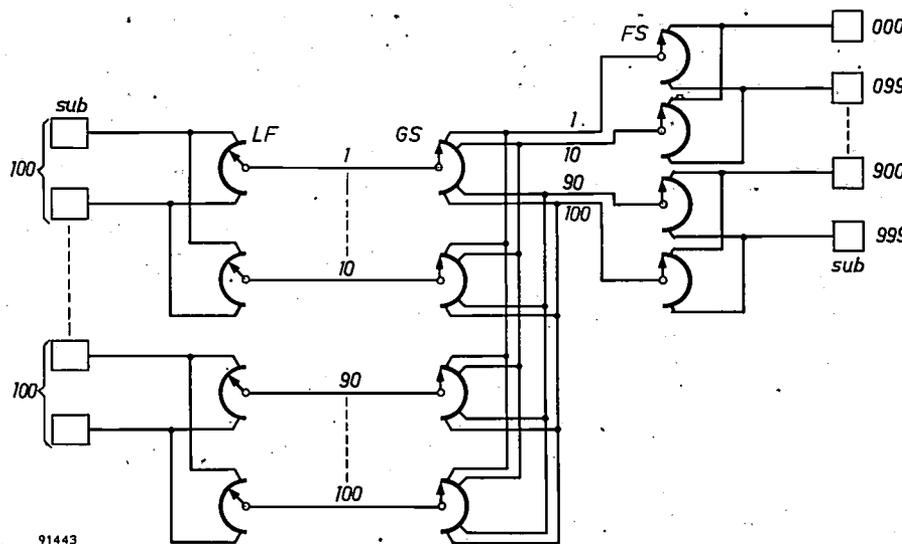


Fig. 1. Elementary scheme of connections in an automatic exchange.

Subscribers' lines (*sub*) to a telephone exchange are generally bunched in groups of 100, being connected up to contact banks of selectors having 100 outlets in order that they can be dialled. These selectors are known as final selectors (*FS*). If the exchange is to handle 100 to 1000 subscribers, group selectors (*GS*) are placed in front of the final selectors, so that an idle final selector can itself be selected as a link to the desired group of 100 subscribers. Where the subscribers number from 1000 to 10 000 a further group-selector stage is necessary in front of the group selectors just mentioned, so that a free group selector can be selected as a link to the desired group of 1000 subscribers.

In order therefore that a subscriber be able to dial, there must be a group selector available for him. Since only a small proportion of subscribers make calls at the same time, it is not desirable to connect each subscriber to the wipers of a group selector. Hence lines from subscribers are connected to a very much smaller number of group selectors via one or more reduction stages. The line finder (*LF*) is an example of a reduction stage. Subscribers' lines are connected to the contacts of the line finders, the wipers of which are connected with the wipers of the group selectors. The number of selectors that is required depends on the number and duration of calls made in the busy hour. For example, a group of 100 subscribers might be provided with 10 group selectors. In this case 10 line finders, each having 100 outlets, would be necessary. When a subscriber takes up his handset the wipers of an idle line finder start passing over the contacts in order to find the caller. For this purpose a relay connected to one of the wipers of the line finder, called the test relay, is put in circuit with the various subscribers' lines one after the other. Immediately the test relay comes into the caller's circuit it operates, thereby bringing the line finder to a halt. In order that calls already in progress between subscribers of the group should not be disturbed, this "testing" is made to take place, not through the two speech conductors, but through a third conductor. Thus not only the two speech conductors link the subscriber to the line finder; there are three conductors, frequently even four. Three or four contacts are therefore necessary for each subscriber. These contacts are placed one above the other, and hence the line finder has three or four layers of contacts and the same number of wipers, the three or four wipers meeting the three or four contacts belonging to a single subscriber. The same applies to group and final selectors, these likewise having three or more layers of contacts.

In the two-motion selector, division into groups is provided by the mechanical construction of the device ( $10 \times 10$  outlets). This is not the case in the uniselector; the grouping of the outlets of a uniselector is effected by electrical interconnection, the result of which is that the dialling of a certain

<sup>2)</sup> The U45a is a further development of the U45 selector also designed by Philips Telecommunicatie Industrie. On the U45, see N. Rodenburg, *Een telefoonsysteem met snelle draaikiezers*, *De Ingenieur* 63, E83-E87, 1951 (No. 50); D. van Hemert and J. Kuin, *Automatische telefonie*, published by Netherlands Post Office, The Hague 1954; J. M. Unk, *A new high-speed uniselector for automatic telephony*, *Comm. News* 12, 69-99, 1951/52.

digit causes all free outlets to the group in question to be "marked" with a suitable voltage. There is no difficulty in altering the number of outlets to be "marked" simultaneously, and hence the division of unselector outlets into groups is very flexible. After a digit has been dialled, an idle outlet (to the final selector or the next group selector, as the case may be) has to be sought amongst the outlets thus "marked". As already stated, it must have been found before the next digit is dialled. The time available is therefore the interval between two trains of impulses, and this is dependent on the temperament of the subscriber. The shortest interdigital pause — the minimum time between two trains of impulses — is approximately 450 msec, 350 msec of which are available for the hunting of an idle outlet. The other 100 msec will be taken up by the release action of a slow relay. If the speed of the selector is not great enough for all the outlets to be swept within 350 msec, then the positioning of the selector must start while the impulses determining the desired group are still being received (i.e. during the return movement of the subscriber's dial), in such a way that the selector arm stands in front of the outlets for that group when the impulse train comes to an end. In that case, only the contacts of the desired group will have to be hunted during the dialling pause. But this would make it necessary for the outlets of a group to be placed next to one another and for the selector arm to have a fixed home position.

If the speed of the selector is great enough for all 100 outlets to be travelled in the 350 msec available, it is no longer necessary that the selector should have a fixed home position or that the positioning should start while the train of impulses is still being received. By the same token, it is not necessary to place all the outlets of a group adjacent to each other; on the contrary, it will be of advantage to distribute them uniformly over the entire arc of the contact bank so that the outlets of various groups occur in the bank one after the other. The selector arm and the wipers will then have, on the average, only a short distance to travel before finding the first idle outlet in the desired group. This shorter distance is about one-tenth of the distance that the wipers would have to travel in the case first considered. There is no need for the selector to return to a home position once the call has finished, and consequently the selector makes only one movement per call. In this way the average finding time and the wear suffered by the selector are very considerably reduced.

The speed necessary for all 100 outlets to be

passed over in approximately 350 msec is  $100/0.35 \approx 300$  outlets per second.

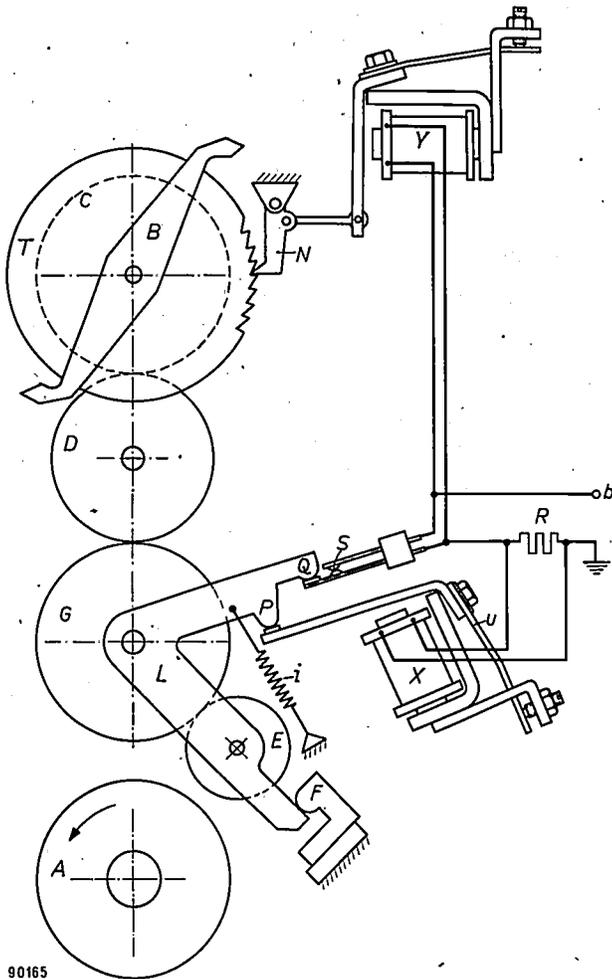
It is desirable that line finders too should work at high speed. Otherwise there is rather a long delay before the calling subscriber hears the dialling tone. Since line finders normally have no home position, and since the wipers of a set of finders will all occupy different random positions, the delay can be cut down by having several idle finders hunting at the same time. The solution is not a satisfactory one, however, for it does not help matters much in the busy hour and the finders suffer increased wear, being set into motion more frequently than is necessary. If a line finder works at high speed it is unnecessary to have more than one engaged in hunting the call. This is so for the U45a selector, which has an *average* finding time of approximately 200 msec, inclusive of the time taken by the relays to operate.

#### Factors limiting the speed of the selector

The main limitation to the speed at which a given selector will work reliably is the need for it to be able to stop in time. The position at which it must come to a halt is normally determined, as stated above, by "marking" a contact with a suitable voltage. Immediately the wiper arrives at this contact, a relay (the test relay) is energized. The tripping of the relay causes the selector to be halted; this must take place so quickly that the wiper does not have time to move on from the "marked" contact. The mechanical shock occasioned by stopping must not be too great, in order that undesirable oscillation of the wipers and the jarring of vulnerable parts of the mechanism be obviated. The shock of stopping depends on the energy of rotation possessed by the selector at that instant; it is therefore of the greatest importance to keep this energy down to a minimum.

We have already said that the U45a selector is designed for a common motor drive: as and when required, the selectors are coupled to a common, continuously running motor. Frequently uniselectors are powered by their own separate motors, in which case they are referred to as motor selectors. In these circumstances the motor is responsible for the greater part of the total energy of rotation. Since this is proportional to the square of the angular velocity and to the moment of inertia of the moving parts, it is obvious that the system of common drive allows of greater speeds, for there is now no need for the motor itself to be stopped, as in the case of the motor selector. In an existing 6-arm motor selector the moment of inertia about the rotor spindle is 2885 g cm<sup>2</sup>. If the selector works

at a speed of 200 contacts per second, the kinetic energy that has to be dissipated when it is brought to a halt is 20.5 millijoule. In a 4-arm U45a selector the moment of inertia about the rotor spindle is 732 g cm<sup>2</sup>, and its kinetic energy at the same speed is 5.8 mJ, or about 3½ times less than the value



90165

Fig. 2. Schematic diagram showing the functioning of the U45a selector. By means of arm *L*, which is actuated by the coupling magnet *X*, pinion *E* can be made to engage with driving wheel *A*. The intermediate member *G* consists of two gears (*y* and *z* in fig. 3) between which there is a friction coupling. *C* and *D* are further gears. *B* is the rotating arm bearing the wipers. *T* and *N* are ratchet wheel and pawl, the latter being actuated by electromagnet *Y*. *Y* is short-circuited via contact *S*. Contact at this point is broken by *Q*, a spur on arm *L*, when *E* is engaged with *A*. The arm *L* has a further spur *P*; *P* is raised, and the arm turned anti-clockwise, by leaf spring *u* acting via the armature of magnet *X* and overcoming the tension of spring *i*, until the other extremity of the arm comes up against stop *F*. *R* is a resistor.

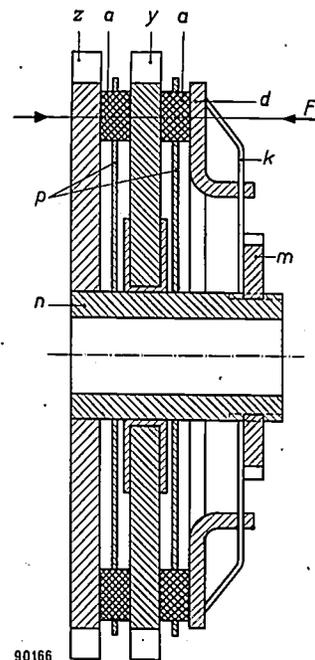
for the motor selector. When the U45a selector works at a rate of 300 contacts per second its kinetic energy amounts to 13 mJ.

Existing selectors with common drive are brought to a halt by withdrawing the driving force once the test relay has tripped, and subsequently applying a braking force. The time available for stopping

the selector is therefore shortened by the length of time necessary for disengaging the coupling, a delay for which the inertia of the coupling is responsible. This severely limits the speed that the selector can attain. In consequence a new type of coupling was developed for the U45a selector, whereby the selector is first brought to a halt and then disengaged by magnetic means from the gear that drives it: a friction coupling between two of the gears allows them to move relative to each other in the interval elapsing between the stopping of the selector and the disengagement of the gears. Hence the inertia of the magnetically-operated coupling plays no part.

### Selector mechanism

The functioning of the selector mechanism will be explained with reference to fig. 2. The rotor consisting of the wipers of the selector, is fixed to gear *C*, which meshes with gear *D*. *G*, the intermediate member, consists of two gears *y* and *z*, which are able to turn relative to each other by virtue of the friction coupling between them (see fig. 3). Gear *z* meshes with *D* and gear *y* with pinion *E*. Pinion *E* is attached to an arm *L*, which by turning causes the former to engage with wheel *A*; this last is fixed to the continuously rotating shaft of the common drive motor.



90166

Fig. 3. Intermediate member (*G* in fig. 2), consisting of gears *y* and *z* and friction blocks *a*, which are kept in place by plates *p*. *F*, the force with which the gears and friction blocks are pressed together, is exerted by spring *k* via tension plate *d*, and can be regulated by means of nut *m*. Gear *z* is fixed with respect to sleeve *n*.

Fixed to the spindle of gear *C* is a ratchet wheel *T*; a pawl *N*, actuated by electromagnet *Y*, engages in the ratchet wheel. When the selector is at rest, the spring *u* of the armature of the coupling magnet *X* presses arm *L* against stop *F*. Electromagnet *Y* is short-circuited via contact *S*. When voltage is applied to terminal *b* of the selector, all that happens in the first place is that coupling magnet *X* pulls its armature over. Arm *L* is now caused by tension spring *i* to turn clockwise, thus coupling pinion *E* with wheel *A*. Spur *Q* on arm *L* breaks contact *S* only after the teeth of *E* and *A* are partially engaged, and only then is electromagnet *Y* energized and the pawl withdrawn from ratchet wheel *T*. Resistor *R* prevents sparking at contact *S* when the latter breaks, and also serves to reduce the reactance of the current path, thus causing electromagnet *Y* to release its armature more quickly when the flow of current is stopped. In consequence of the fact that *E* and *A* are already partially engaged, and that therefore gear *y* in intermediate member *G* is already turning, there is slipping of the friction coupling in *G* and a torque is exerted on gear *z*, which meshes with *D*. Thus the wipers start to turn immediately the pawl retracts. This sequence of events means that the friction coupling absorbs the shock resulting from the meeting of the wheel *A* and the pinion *E*, and hence there is no question of the wipers undergoing excessive acceleration.

The torque transferred by the friction coupling (see fig. 3) is proportional to the coefficient of friction between the blocks and the steel gears. For the frictional material employed (leather blocks impregnated with a silicon oil), the static coefficient of friction is less than the kinetic, and this latter increases with the amount of slip. The torque is therefore strongest at the moment the pawl retracts, for that is the time when the greatest amount of slip occurs. In consequence, the selector is brought up to its normal speed in the shortest possible time. While it is thus being speeded up, the amount of slip between gears *y* and *z* decreases, with the result that the torque decreases and the acceleration of the wipers proceeds very evenly. Because the coefficient of friction for the material employed decreases with increasing temperature, it will never heat up to excessive temperatures and, at the normal rate of rotation, the friction coupling will allow continuous slipping to take place without exhibiting perceptible wear or any tendency to jam.

When the selector has to be stopped, the flow of current through stopping magnet *Y* and coupling magnet *X* is broken by the test relay. Release by the stopping magnet of its armature causes the

pawl to arrest the ratchet wheel (and hence the wipers); spring *u* causes arm *L* bearing pinion *E* to turn anticlockwise, thus disengaging *E* from *A*. In the interval elapsing between the stopping of the wipers and the disengagement of the selector there is slipping of the friction coupling. The torque that continues to be exerted on *C* and hence on ratchet wheel *T* serves to counteract bounce. Since the torque increases with the amount of slip, the braking torque on the rotor during its small backward movement following bounce is on the average greater than the driving torque exerted in forward movement. The absorption of the kinetic energy of the rotor by friction between wipers and contacts and the elastic deformation occasioned by collision between pawl and ratchet also contribute to the reduction of bounce. At each collision about half of the energy is dissipated. All these effects result in the rotor coming practically to a halt after the second bounce.

The movement of the wipers in relation to the contacts when the selector stops is shown in fig. 4. It is quite clear from the graph that the wipers do not fly past the contacts when the ratchet wheel bounces back from the pawl; indeed, there is a considerable margin of safety.

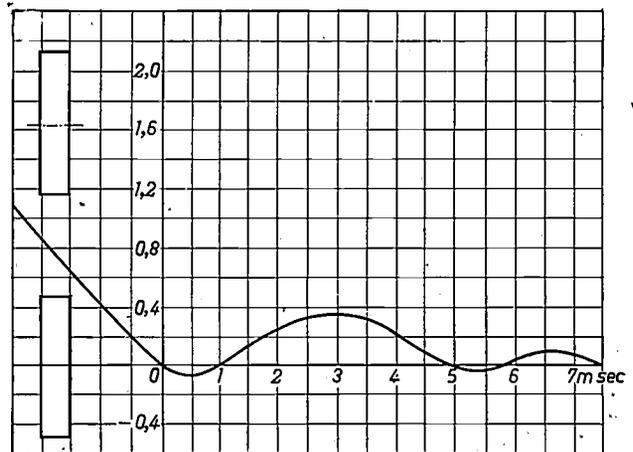


Fig. 4. Displacement/time diagram showing the stopping of the selector (which has a speed of rotation of 300 contacts per second). The width and position of two successive contacts are indicated on the left.

The U45a selector as described here has a separate ratchet wheel for the stopping mechanism. In other designs use is made of the final gearwheel on the wiper spindle. In this case, however, the tooth flank that comes up against the pawl is not perpendicular to the line *OO'* (see fig. 5a). The consequence is that there is a fairly strong force tending to push the pawl out of the teeth when the selector

is, halted, and a force  $F_a$  has to be exerted on the pawl in order to prevent this happening. Where an independent ratchet wheel is employed its teeth can be cut in such a way that the flank is perpendicular to  $OO'$  (see fig. 5b), so that there is not the slightest tendency for the pawl to be thrown out

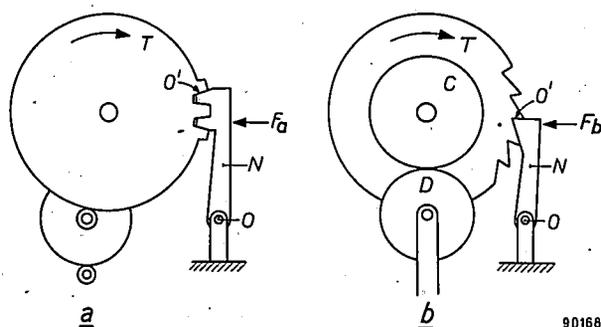


Fig. 5. a) Gearwheel  $T$  and pawl  $N$  for stopping the selector. The flank of the tooth at  $O'$  is not perpendicular to a line drawn from  $O$  to  $O'$ .  
b) Ratchet wheel  $T$  and pawl  $N$  for stopping the selector. Here the flank of the tooth at  $O'$  is perpendicular to  $OO'$ .

of the ratchet. A weak force ( $F_b$ ) exerted by the armature of the stopping magnet on the pawl suffices to prevent its ejection, and this makes it possible for the armature of the magnet to be of quite light construction. Accordingly, the power required for the stopping magnet in the U45a selector is only 6 watts. This has the advantage that the current through the magnet is weak, so weak that if a fault should develop and the winding remain in circuit, it would not overheat and therefore would not burn out. Despite the weak energizing current, the pressure of the spring on the armature is such that the pawl halts the ratchet wheel 1 msec after the flow of current through the stopping magnet has been broken, reaching the apex of the ratchet trough after 1.2 msec. Tests on the selector in a special circuit (in place of the test relay circuit) have proved it to be possible to stop the selector at any desired contact from a speed of 700 contacts per second.

Apart from the shape of the teeth, the diameter of the ratchet wheel is of great importance. For a given total kinetic energy, the shock (change of momentum) when the ratchet wheel collides with the pawl decreases proportionally as the diameter of the wheel increases. On the other hand, the moment of inertia of the wheel increases as the fourth power of its diameter. For a given speed of rotation, therefore, the kinetic energy of the ratchet wheel, and hence the total kinetic energy, increases rapidly with the diameter of the wheel. There is a

certain diameter of the wheel for which the shock of collision is at a minimum. For a U45a selector with 102 outlets and 6 wipers it is round about 70 mm, and this is the diameter that has been chosen for the ratchet wheel. Life tests have demonstrated that ratchet and pawl suffer no marked wear after having acted  $10^7$  times; this is equivalent to  $10^7$  calls, or a call every five minutes for a century.

In general, the total kinetic energy to be dissipated per call gives a measure for wear on the selector and for contact noise. The energy to be dissipated per call in a U45a selector working at a speed of 300 contacts per second is approximately 13 mJ. The equivalent figure for the earlier-mentioned motor selector working at a speed of 200 contacts per second (this involving two halts per call) is approximately 41 mJ in all. In the two-motion selector working at a speed of 10 steps per second during vertical movement and 30 steps per second during rotary movement, the total energy to be dissipated per call amounts to approximately 400 mJ.

#### Constructional details of the selector

The ratchet wheel in the standard version of the U45a selector is provided with 102 teeth, each tooth corresponding to an outlet. The selector therefore has 102 outlets. The contacts of these outlets (there may be 4 or 6 contacts for each outlet, arranged in 4 or 6 layers — see account under fig. 1) together form the contact bank. In order to make it possible to extract the wiper mechanism, these contacts are not distributed over the full circumference of a circle; instead, the 102 contacts of each

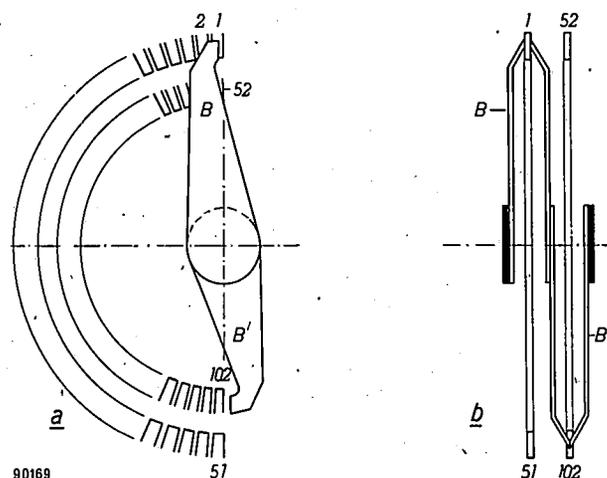


Fig. 6. Diagram to show distribution of 102 contacts over two arc-shaped rows. In (a) for the sake of clarity, the arc containing contacts 52-102 is drawn with a smaller radius than the arc containing contacts 1-51 and wiper  $B'$  is drawn smaller than wiper  $B$ . The true proportions are given in the side-view (b).

layer are divided into two 180° arcs of 51 contacts each. The two arcs are fixed side by side. The selector has two wipers for each layer; the two wipers are diametrically opposed and each travels over an arc in its turn. The arrangement is shown in *fig. 6*. For the sake of clarity, *fig. 6a* has been drawn in such a way that the radius of one arc of contacts, and the wiper for that arc, appear longer than the radius and wiper of the other arc, although in reality they are identical and situated adjacent to each other. Wiper *B* travels over contacts 1-51; as soon as it leaves contact 51, wiper *B'* arrives at contact 52 and passes successively over the contacts up to 102.

If desired, the number of outlets on the selector can be reduced to 51. There is then only one arc of 51 contacts per layer, these 51 contacts being distributed over 180°. In this case, the single arc is travelled by a single wiper with two ends, so that the wiper reaches contact 1 immediately after its other end leaves contact 51 (see *fig. 7*). The selectors can be made with any desired number of layers up to 16. Since it is not economic to keep

a variety of kinds in stock, two standard versions are produced, one with 4 and one with 8 layers.

A selector with 4 wipers and 102 outlets appears in *fig. 8*; *fig. 9* shows the wipers complete with the driving mechanism. This unit can be extracted

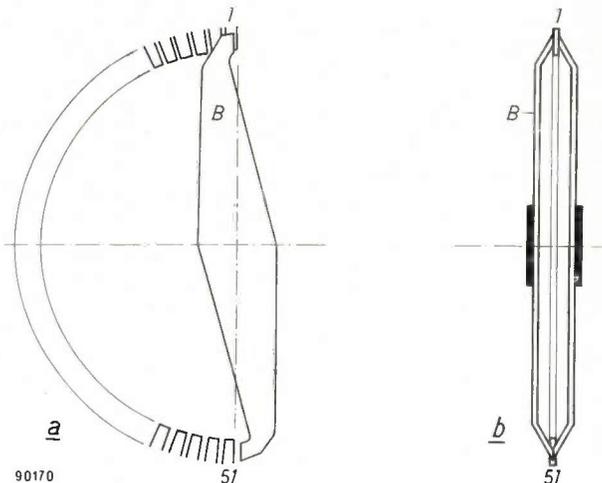


Fig. 7. a) Simplified diagram of selector having 51 contacts and double wiper arm *B*. b) Side-view of the same.

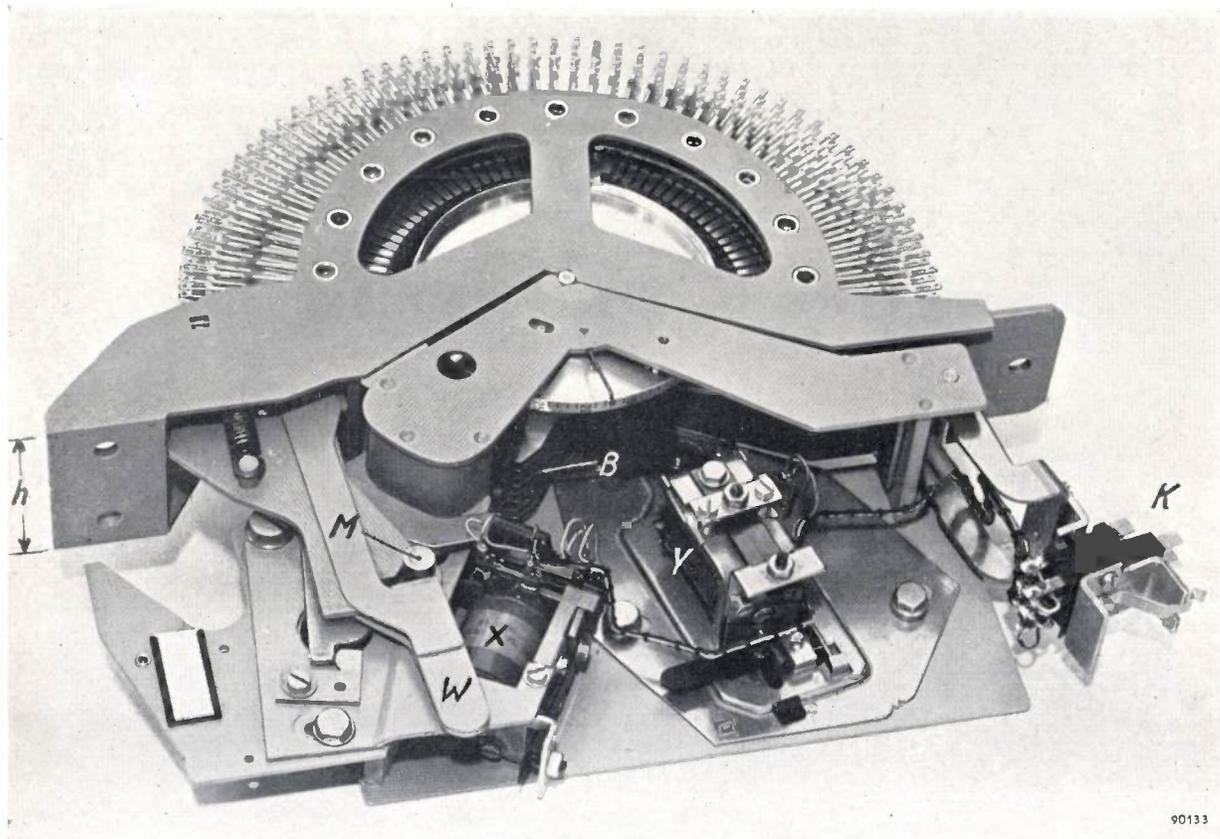


Fig. 8. U45a, a high-speed uniselector with four wipers and 102 outlets. *X* and *Y* are the coupling and stopping magnets respectively. *W* is a lever which, by means of a pin *M*, retains the wiper mechanism in the contact bank. *K*: knife contacts for electrical connections to magnets *X* and *Y* and wipers *B*. The overall breadth of the selector is  $h = 49$  mm.

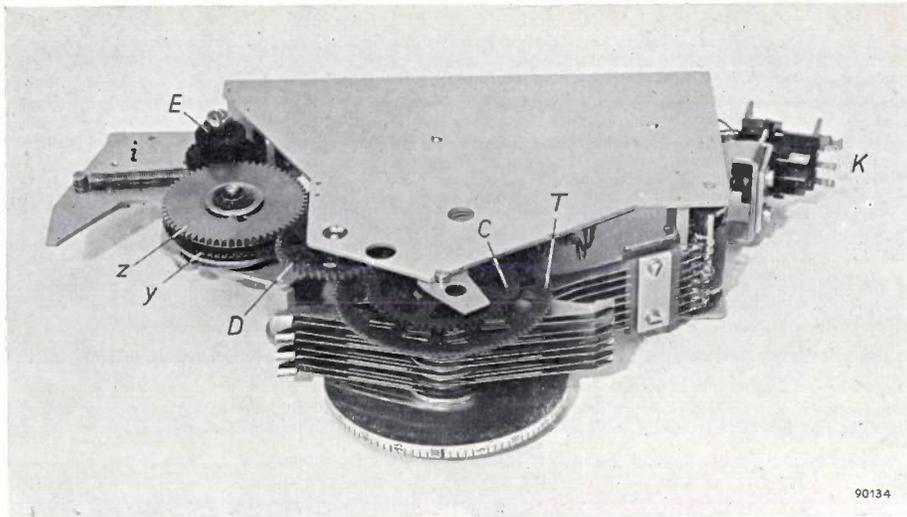


Fig. 9. Wiper mechanism with four wipers. *y* and *z* are gears of the intermediate member, *E* is the coupling pinion, *C* and *D* are gears, *T* is the ratchet wheel, *i* the tension spring (see fig. 2) and *K* the knife contacts.

from the contact bank in a simple fashion, even while in operation. First the knife contacts *K* are released (see fig. 8) by turning the arm on which they are mounted. Lever *W* is then moved to the left,

thereby freeing the pin *M* and making it possible to take the whole wiper unit out of the contact bank. Lever *W* is provided with a spring whereby a force is exerted on pin *M*, and whereby the sta-

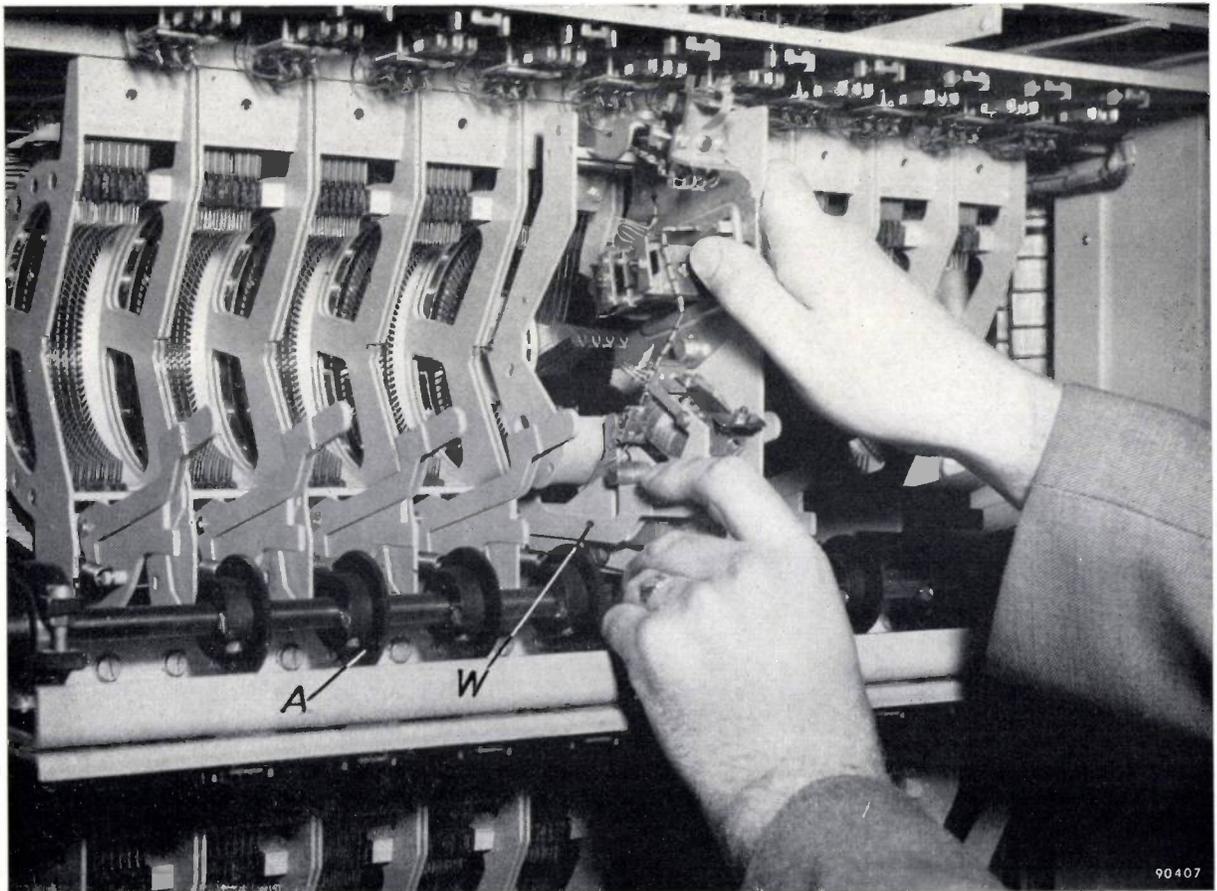


Fig. 10. Inserting a four-wiper selector in a rack. *A* is the continuously rotating shaft of the common drive. *W* is the lever shown in fig. 8.

tionary spindle on which the wipers and ratchet wheel rotate is pressed into two V-shaped recesses in the end-plates of the contact bank, good centering thus being ensured. Provision has been made for easy adjustment of the positions of the coupling and stopping magnets on the mounting-plate of the wiper unit. This is necessary in order that, when the wiper unit is inserted in a contact bank for the first time, the coupling magnet may be correctly placed in relation to the driving shaft and the stopping magnet may be so positioned that the wipers stop in the right place on the contacts.

The selectors are mounted in the rack side by side in groups, the contact banks being arranged vertically (*fig. 10*). The horizontal driving shaft for each group is itself driven by a chain common to the whole rack. Chain transmission has the advantage of enabling the racks to be fairly light in construction, for it will not be affected by vibrations in the building or by slight departures from exactness in the mounting of the horizontal driving shafts.

The chains are driven off a horizontal shaft along the foot of each row of racks. Should any of the chains be too heavily loaded, it is automatically disengaged from the main driving shaft. It is possible to operate these safety couplings by hand.

---

**Summary.** The importance of high speed in automatic telephone selectors is discussed with reference to several examples of such devices. It is demonstrated that, where one-motion selectors (uniselectors) are employed, it is advantageous to make them work at a speed such that at least 300 outlets can be swept per second. There is then no need for the selector to have a home position and, furthermore, it is possible for outlets of groups to be distributed uniformly over the whole of the contact bank. This offers the great advantage that only one movement is made per call, the average distance travelled per call being one-tenth of that which would have to be travelled if there were a home position. In this way the average finding time and the wear suffered by the selector are greatly reduced. In the U45a rotary selector developed by Philips Telecommunicatie Industrie, the speed mentioned above is attained by employing a common driving mechanism with gear transmission and a new kind of coupling, whereby the selector is first brought to a halt and then disengaged. A friction coupling between the gears allows them to move relative to each other in the time intervening between stopping and the disengagement of the drive.

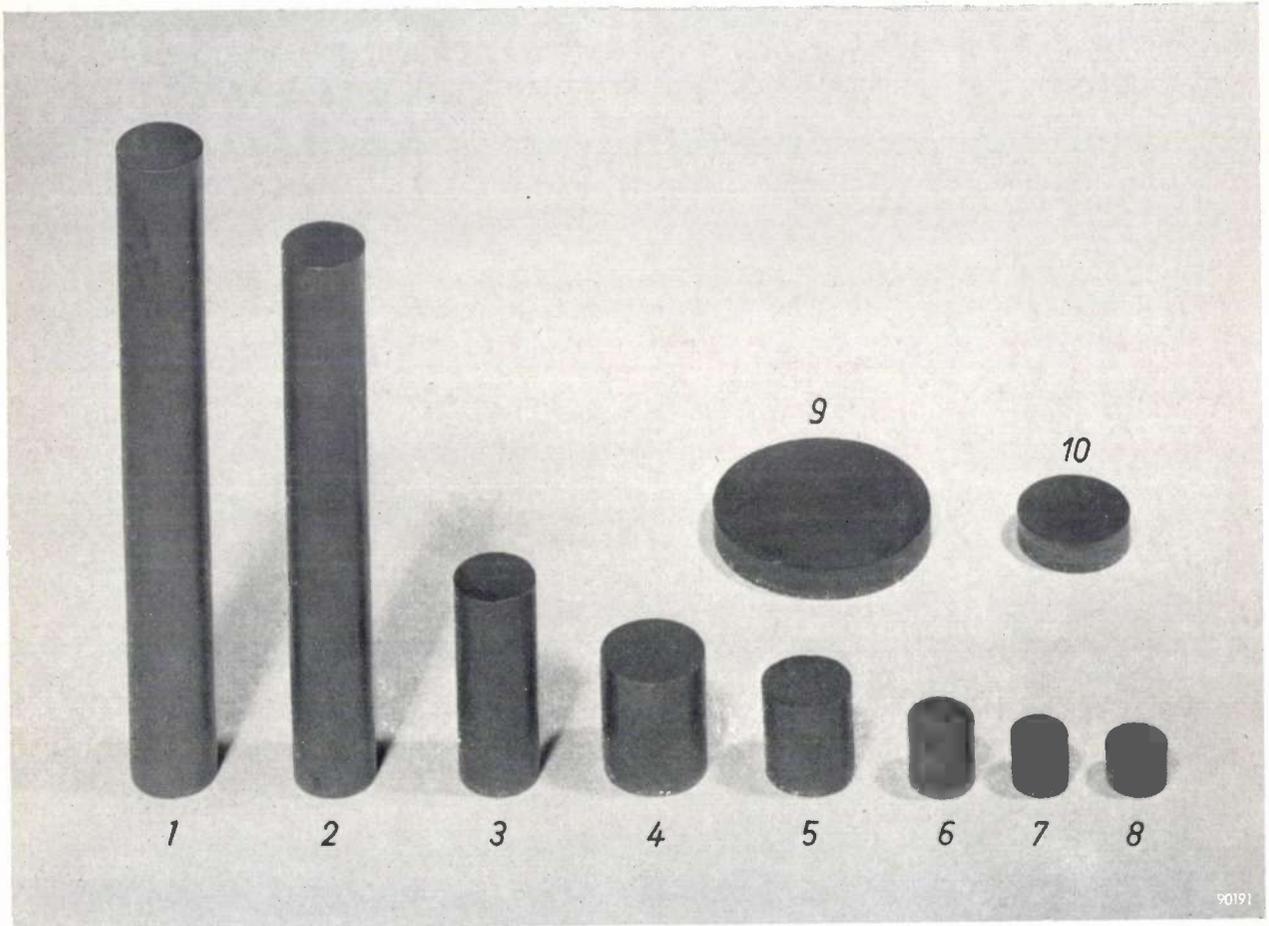
## THE EVOLUTION OF THE PERMANENT MAGNET: A BRIEF REVIEW

621.318.2

In evaluating a material for making permanent magnets it is the magnetic properties that are of primary interest. Many other properties play a secondary part, such as workability, resistance to oxidation, the stability of the magnetic properties and, finally, the availability of raw materials and the price of the magnet.

that the operating point on the demagnetization curve is precisely that point at which the  $BH$  product has its maximum<sup>1)</sup>  $(BH)_{\max}$ .

The photograph will serve to illustrate the evolution of the permanent magnet from the point of view of the  $(BH)_{\max}$  value. It shows a series of magnets made of different materials, each magnet



The quality of a magnetic material as regards its magnetic properties can be characterized fairly well by the maximum value, as occurring in the demagnetization curve, of the product of induction  $B$  and the corresponding field strength  $H$ . The product  $B \times H$  defines the magnetic energy obtainable per unit volume of the magnetic material when placed in a magnetic circuit to give a field in an air gap. The greater  $(BH)_{\max}$ , the smaller are the dimensions with which the magnet can be made in order that it should possess a specified total magnetic energy (i.e. create a specified magnetic energy in an air gap). It is then, of course, necessary to design the magnet and the magnetic circuit in such a way

being of such a volume that they would all give the same total magnetic energy in a (hypothetical) air gap. Each magnet has in fact a cross-section and a length such that, when used at their  $(BH)_{\max}$  point, all magnets supply the same flux ( $B \times$  area of cross-section) and the same magnetomotive force ( $H \times$  length). A well-designed loudspeaker assembly served as basis for the choice of these values of flux and magnetomotive force. Needless to say, the new magnets designed to give those values would not fit into the magnetic circuit of the original assembly and, were the original circuit adapted to

<sup>1)</sup> For the design of permanent magnet circuits, see for example A. Th. van Urk, Philips tech. Rev. 5, 29-35, 1940.

the new magnets, the complication of a different stray field would arise (see the article cited in footnote<sup>1</sup>). If we leave this correction aside, however, the photograph gives a striking picture of the progress that the development of new materials has rendered possible, and also of the effect of their development on the proportions of magnets. Some details of the materials used for the magnets in the photograph will now be given (see also the table below).

On the extreme left is a magnet made of the oldest kind of magnetic steel, namely carbon steel having a  $(BH)_{\max}$  of  $0.26 \times 10^6$  gauss.oersted<sup>2</sup>). Next to it is a magnet of tungsten steel (almost identical

years passed before it came into general use; the then youthful radio industry, in which stronger yet smaller and lighter magnets were in demand for loudspeakers, was largely responsible for its introduction. The fourth magnet in the photograph is of Mishima alloy, a magnetic alloy based on nickel-aluminium-iron, which was developed in Japan in 1933 and which has the somewhat higher  $(BH)_{\max}$  of  $1.05 \times 10^6$ . The price of such a magnet is about a third of that of an equivalent cobalt-steel magnet. Experiments on the addition of other metals to the alloy, mainly cobalt, titanium and copper, and on strictly controlled heat treatment, were successful in further improving the properties of Mishima

Table. Further data on the materials from which the permanent magnets shown in the photograph were prepared. Magnet 7a does not appear in the photo, but its material is discussed in the text.

No. in photo	Year	Designation	Magnetic properties					Dimensions of magnets		
			$B_{\text{rem}}$ (gauss)	$H_{\text{coerc}}$ (oersted)	At optimum working point			Diameter (mm)	Length (mm)	Weight (gm)
					$B$ (gauss)	$H$ (oersted)	$(BH)_{\max}$ ( $10^6$ gauss $\times$ oersted)			
1	1880	Carbon steel	10000	50	5400	48	0.26	19	136	294
2	1900	Tungsten steel	11000	70	6000	56	0.34	18	115	225
3	1917	Cobalt steel 35% (Honda)	9200	240	6000	150	0.90	18	43	83
4	1938	Ni-Al alloy (Mishima)	6100	480	4050	260	1.05	22	25	69
5	1936	"Ticonal" II	6300	780	5300	340	1.80	19	19	37
6	1937	"Ticonal" G (anisotropic)	13000	580	10000	500	5	14	13	14.5
7	1949	"Ticonal" GG *) (orient. cryst. struct.)	14500	720	13000	640	8.3	12	10	8.5
7a	1953	"Ticonal" X (great coercivity)	9000	1250	5000	900	4.5	19.5	7.2	16
8	1956	"Ticonal" XX *)	11800	1315	9200	1200	11.0	14.7	6	7
9	1952	Ferroxdure I	2000	1800	1000	900	0.9	44	7.2	53
10	1954	Ferroxdure II	3750	1300	2640	1100	2.9	27	6	17.2

\*) Material prepared on laboratory scale.

with chrome steel, both being developed before the first world war) having a  $(BH)_{\max}$  of  $0.36 \times 10^6$ . In 1917 Honda steel appeared; it contained cobalt and had a much higher  $(BH)_{\max}$  value, viz.  $0.9 \times 10^6$  for the magnet shown, which is of steel with 35% cobalt. This steel differed to such an extent from the types then known, both in the attainable magnet dimensions and in price (cobalt is about one hundred times dearer than iron), that

alloy. In this way "Ticonal" II was developed, this having a  $(BH)_{\max}$  of  $1.8 \times 10^6$ ; it was taken into production about 1936.

In 1938 an important step in the evolution of magnet steels was made in the Philips laboratories, when Jonas introduced heat treatment in a magnetic field<sup>3</sup>) for certain well-defined groups of

<sup>2</sup>) The values of the  $(BH)_{\max}$  product are given in gauss oersted in accordance with established practice, rather than in rationalized Giorgi units commonly used in this Review. In the rest of the article units will be omitted when  $(BH)_{\max}$  values are quoted.

<sup>3</sup>) Netherlands Patent No. 53734; B. Jonas and H. J. Meerkamp van Embden, Philips tech. Rev. 6, 8-11, 1941. A short time previously Oliver and Shedden (Nature 142, 209, 1938) had carried out similar tests on magnetic alloys already in use, producing a very slight improvement in their properties. Owing to an incorrect interpretation of the phenomenon, these investigators did not at the time continue their observations.

alloys containing nickel, cobalt, aluminium, copper and iron (and in certain cases other elements such as titanium). In this way the  $(BH)_{\max}$  value was raised to over  $5 \times 10^6$  (in "Ticonal" G, the sixth magnet from the left). So important was this step that, even today, about a half of the world production of magnets is carried out according to the process.

In the years following 1933 much work was devoted to the problems of magnetism all over the world. Understanding of magnetic phenomena was deepened by investigators such as Néel, Snoek, Kersten and many others, and various new materials were discovered, such as the iron-cobalt-molybdenum alloys of Sykes and Köster, the cunife and cunife of Dannöhl et al., the cobalt-platinum alloys of Jellinghaus, the "Vicalloy" of Nesbitt et al.; the powdered magnetic materials of Néel and the "Bismanol" (bismuth-manganese alloy) of Guillaud and the Naval Ordnance Laboratory in Washington. Considerations of cost price or practical difficulties, or both, prevented these materials being applied except in rare cases, however.

It was not until 1949 that  $(BH)_{\max}$  was further stepped up. The normal "Ticonal" alloy was so diffused from the melt in a special manner so that the crystals of the material obtained had a certain orientation. By subjecting magnets of this kind to the heat treatment previously devised<sup>4)</sup>, very high values of  $(BH)_{\max}$ , rising to  $8.4 \times 10^6$ , were attained in the laboratory. It was found possible to produce these materials on a larger scale with  $(BH)_{\max}$  values of  $6.5$  to  $7.5 \times 10^6$ .

In the meantime efforts have continued to raise  $(BH)_{\max}$  still higher. In recent years the record has been held by the American General Electric Co. with a cobalt-platinum alloy. The alloy had been described by Jellinghaus much earlier; however, in 1955, by subjecting it to a heat treatment which ordered the crystal lattice, Martin and Geisler<sup>5)</sup> raised its  $(BH)_{\max}$  value to  $9 \times 10^6$ . Its price naturally prohibits the extensive employment of this material.

The record has now been broken again in the

Philips laboratories at Eindhoven. Earlier, a method had been discovered of increasing the coercive force of "Ticonal" alloys. The method consists of the addition of titanium and the raising of the cobalt content<sup>6)</sup>, combined with a special kind of heat treatment<sup>7)</sup>. It is possible thus to prepare materials ("Ticonal" X) with a coercivity  $H_{\text{coerc}}$  of 1300 Oe and a  $(BH)_{\max}$  of  $4.5$  to  $5 \times 10^6$ . In the course of attempts to improve this material still further by solidifying it in the desired orientation, Luteijn and De Vos succeeded a short time ago in obtaining, in the laboratory, magnets having crystal orientation with a  $(BH)_{\max}$  of  $11 \times 10^6$ <sup>8)</sup>; for this purpose they worked with very pure constituents, fusing them in an inert atmosphere. In the row of magnets shown in our photograph, the one on the extreme right, the smallest of all, is made of this material.

In addition, the photograph shows two permanent magnets made of ferroxdure, a ceramic (non-metallic) material; these two have the same flux and same magnetomotive force, and hence the same magnetic energy, as the other magnets appearing in the picture. Although the development of ferroxdure<sup>9)</sup> was governed by quite other considerations than the attainment of extremely high  $(BH)_{\max}$  values, it is instructive to include these two magnets in the series; the relatively low remanence but extraordinarily high coercive force of ferroxdure result in the rod form of the magnet we started with becoming a disc. The  $(BH)_{\max}$  of ferroxdure I lies in the vicinity of  $0.9 \times 10^6$ ; in the meantime this value has been raised to above  $3 \times 10^6$  (maximum  $3.8 \times 10^6$ ) for ferroxdure II, this having been achieved by aligning the particles of the basic material in a magnetic field prior to sintering. It may be remarked that both the great coercive force of this material, and its non-metallic character, has resulted in a wide range of special applications.

<sup>6)</sup> Netherlands Patent Nos. 70092 and 70424.

<sup>7)</sup> Netherlands Patent applied for.

<sup>8)</sup> A. I. Luteijn and K. J. de Vos, Philips Res. Rep. 11, 489-490, 1956.

<sup>9)</sup> J. J. Went, G. W. Rathenau, E. W. Gorter and G. W. van Oosterhout, Philips tech. Rev. 13, 194-208, 1951/52; A. L. Stuijts, G. W. Rathenau and G. H. Weber, Philips tech. Rev. 16, 141-147, 1954/55.

<sup>4)</sup> Netherlands Patent No. 71925.

<sup>5)</sup> D. L. Martin and A. H. Geisler, J. appl. Phys. 24, 498, 1953.

## A LOW-NOISE KLYSTRON WITH HIGH POWER OUTPUT

by R. A. LA PLANTE \*) and G. A. ESPERSEN \*).

621.373.423

*Among the tubes used for generating a continuous high power at high frequencies the velocity-modulation valve or klystron occupies an important place. In this article the authors describe an investigation on the noise of these tubes. The principle cause was shown to be situated in the tuning mechanism. A non-tunable tube, constructed as a result of these experiments, shows a very low noise-level and has also some other favourable properties.*

Velocity-modulation valves or klystrons have already been dealt with in two previous articles in this Review<sup>1)</sup>. For a description of the working principles we refer therefore to these papers. In the second of them construction details were given of some tubes developed for wavelengths between 15 and 3 cm (frequencies between 2000 and 10000 Mc/s). In this article we will describe some recent investigations on the subject of klystrons, more especially with regard to the causes of noise in these tubes. We will also describe the construction of a klystron that has been developed as a corollary to these investigations.

### Construction of a 3 cm klystron

Fig. 1 shows a simplified cross-section of a 3 cm klystron developed in Eindhoven some years ago. The construction is basically the same as that of the tubes described in the second article mentioned under <sup>1)</sup>, the principle difference being that the output cavity resonator is not coupled to the waveguide by a loop but by a gap in the resonator wall, the gap being connected to a waveguide flange via a tapered waveguide. The vacuum seal is obtained with a mica window of thickness 50 microns. The tube has an L cathode<sup>2)</sup> which, even at the high emission density required, has a useful life of more than 1000 hours. This tube delivers a power output of 200 W. at an anode voltage of 8800 V, and has an efficiency of 13%.

\*) Philips Laboratories, Irvington-on-Hudson, N.Y., U.S.A.

<sup>1)</sup> F. M. Penning, Velocity-modulation valves, Philips tech. Rev. 8, 214-224, 1946, and B. B. van Iperen, Velocity-modulation valves for 100 to 1000 watts continuous output, Philips tech. Rev. 13, 209-222, 1951/52.

<sup>2)</sup> H. J. Lemmens, M. J. Jansen and R. Loosjes, A new thermionic cathode for heavy loads, Philips tech. Rev. 11, 341-350, 1949/50.

### Noise of a klystron

As with all oscillators, the currents and voltages generated in a klystron are not purely sinusoidal but are subject to variations both in amplitude and phase. This also applies, therefore, to the electromagnetic field in the waveguide coupled to the tube. In many cases this deviation from a purely sinusoidal waveform manifests itself as noise, which restricts the usefulness of the tube. For instance, noise

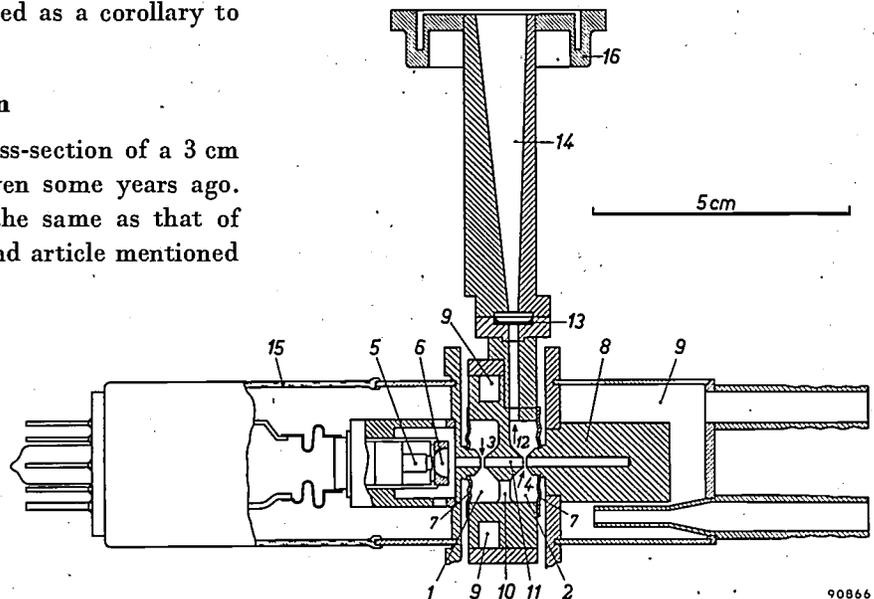


Fig. 1. Cross-section of a tunable klystron for a wavelength of 3 cm. Various details, including the tuning mechanism, have been omitted. 1 modulating cavity, 2 output cavity, 3 modulating gap, 4 inductor gap, 5 cathode, 6 focusing electrode ("grid"), 7 diaphragms, 8 collector, 9 cooling jacket, 10 feedback hole, 11 drift tube, 12 coupling gap, 13 mica window, 14 tapered waveguide, 15 glass envelope, 16 waveguide flange.

in klystrons used in beam transmitters reduces the range of the transmitter and may also limit the number of non-interfering channels permissible in a given frequency band<sup>3)</sup>.

In the ideal case of a completely noise-free signal the field at every point in the waveguide would vary as a function of time according to  $A \cos \omega_0 t$ .

<sup>3)</sup> See C. Ducot, Beam transmitters with double frequency modulation, Philips tech. Rev. 17, 317-327, 1955/56.

For variations in amplitude  $A$  and in phase  $\omega_0 t$  the formula for the field may be written as:

$$A \{1 + a(t)\} \cos \{\omega_0 t + p(t)\},$$

where  $a(t)$  and  $p(t)$  are random time functions.

It has been established that the noise is mainly due to phase variations and that  $\sigma(t)$  is always very much smaller than unity.

Experiments, which we shall describe in this paper, have shown that the noise is predominantly due to microphony, to which tubes of this construction are particularly sensitive because of the presence of the two diaphragms (7). The extremely small vibrations of these diaphragms give rise to slight variations in the tuning frequency of both cavity resonators, resulting in phase variations and hence frequency variations of the output signal.

Another cause of noise, viz. fluctuations in the applied voltages, especially the anode voltage, can be kept very small by using adequately stabilized voltages.

table attenuator<sup>4)</sup>,  $At_1$ , to a crystal detector  $D_1$ . Since the signal voltage appearing on this detector is very small, the output of the rectifier is proportional to the square of the signal voltage and therefore proportional to the output power of the klystron. This D.C. voltage, which contains a small A.C. component due to amplitude variations in the signal, can be read from meter  $M_1$ .

The signal from  $DC_2$  is fed via a second adjustable attenuator,  $At_2$ , to a discriminator; this consists of a slightly detuned transmission cavity,  $C$ , to which a crystal detector,  $D_2$ , is coupled. The D.C. voltage from this detector can be read from meter  $M_2$ . As the discriminator detects fluctuations in amplitude as well as in frequency, the D.C. voltage output of this channel contains two A.C. components, one proportional to amplitude variations and the other to frequency variations in the klystron signal (there are also some small interaction terms present). When the D.C. outputs of both detectors are equal, the A.C. components in each output are also equal

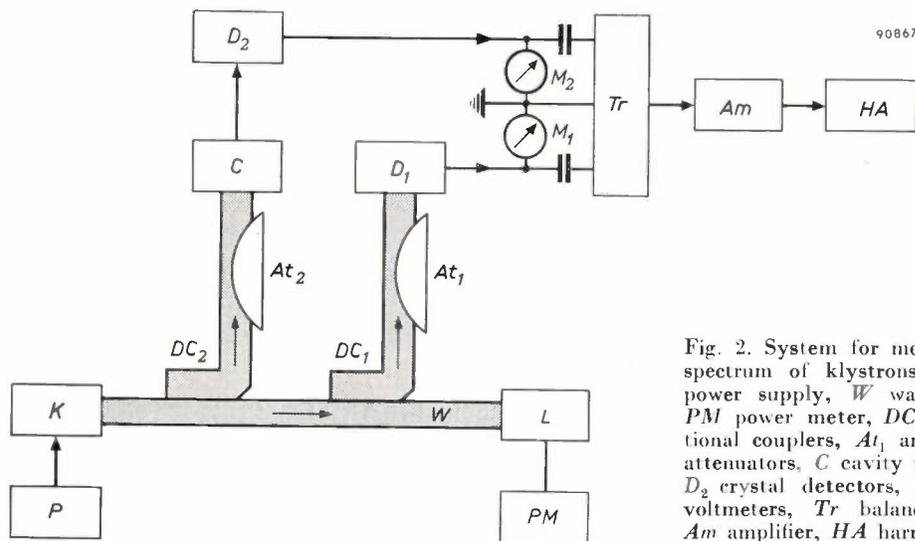


Fig. 2. System for measuring the noise spectrum of klystrons.  $K$  klystron,  $P$  power supply,  $W$  waveguide,  $L$  load,  $PM$  power meter,  $DC_1$  and  $DC_2$  directional couplers,  $At_1$  and  $At_2$  adjustable attenuators,  $C$  cavity resonator,  $D_1$  and  $D_2$  crystal detectors,  $M_1$  and  $M_2$  D.C. voltmeters,  $Tr$  balancing transformer,  $Am$  amplifier,  $HA$  harmonic analyser.

### Noise measurements

Fig. 2 shows a block diagram of the system used for measuring the noise of klystrons. The tube  $K$  under investigation operates into a waveguide  $W$  terminated by a matched load  $L$ . Ignoring the slight losses in the waveguide, the output power of the klystron can be ascertained by measuring the power converted into heat in this load.

Two samples of the klystron output are "taken" from the waveguide  $W$  and directed into separate channels by two directional couplers<sup>4)</sup>  $DC_1$  and  $DC_2$ .

The signal supplied by  $DC_1$  is fed via an adjust-

in so far as they are due to amplitude variations. The A.C. outputs from  $D_1$  and  $D_2$  are applied to a push-pull-transformer  $Tr$  in such a way that the voltage obtained in the secondary is mainly caused by frequency variations. This voltage is amplified in an audio amplifier  $Am$  and analysed by a harmonic analyser  $HA$ . From this voltage we can calculate the power spectrum of the klystron, the width of this spectrum being a measure for the noise.

A total of four klystrons of the type shown in fig. 1 were analysed in this way, with the results shown in fig. 3. Here we plot the power against the frequency, expressed in W per c/s. A measure for

<sup>4)</sup> See A. E. Pannenberg, A measuring arrangement for waveguides, Philips tech. Rev. 12, 15-24, 1950/51.

the width of these spectra is the so-called standard deviation

$$\gamma = \sqrt{\frac{\int_0^{\infty} x^2 S dx}{\int_0^{\infty} S dx}}$$

For the curves shown in fig. 3 the value of  $\gamma$  lies between 3.3 and 12 kc/s.

The fact that the noise is mainly due to the microphony arising from the diaphragms was plainly demonstrated by the following experiment. After the spectrum of one of the tubes had been recorded, the diaphragms and tuner were cast in plaster of Paris, and the spectrum again recorded. The spectra before and after this treatment are represented as curves *a* and *b* in fig. 4. The root mean square deviation was reduced in this way from 5.3 to 1.9 kc/s. The plaster of Paris was then removed and the tube was cast in plastic, after which the spectrum was again recorded. The result is shown as curve *c* in fig. 4. The root mean square deviation was now found to be only 200 c/s.

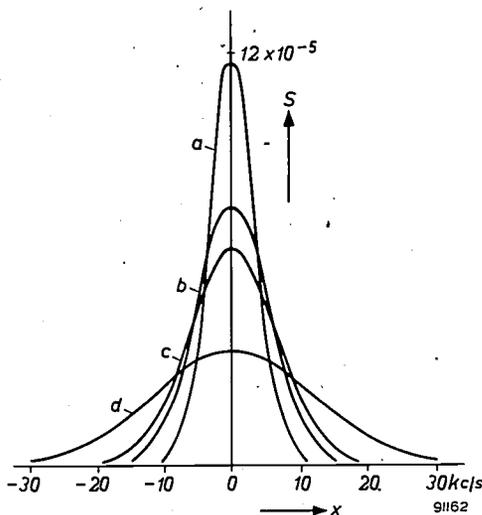


Fig. 3. Spectrum of four tunable klystrons, recorded with the layout shown in fig. 2. The power  $S$ , expressed in W per c/s, is set out along the ordinate;  $x$  is the frequency deviation from the central frequency of the spectrum.

**Construction of a low-noise tube**

The main source of the noise having been determined, various methods were considered for reducing microphony. Firstly attempts were made to build a more robust tuner to hold the diaphragms more firmly (see fig. 8), but the results showed very little improvement, the r.m.s. deviation of the tubes so constructed being of the order of 2 kc/s. As a low noise level was so important that the tuneability of the tube could if necessary be sacrificed for this

purpose, it was furthermore considered to cast the tubes in plastic, after first tuning them to a specified frequency. It would also be possible to make tubes with only one diaphragm, i.e. with one of the cavities tuned to a fixed frequency, and finally a tube might be designed with no diaphragms at all.

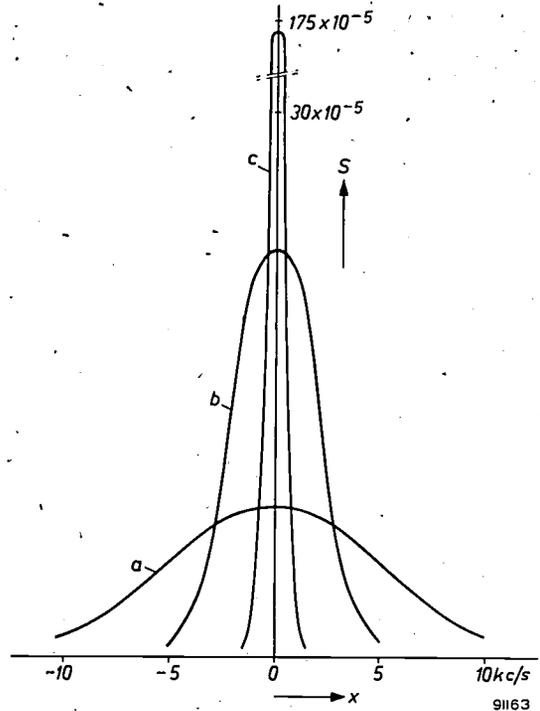


Fig. 4. Spectrum of a tunable klystron, *a* in normal state, *b* after casting in plaster of Paris, *c* after casting in plastic.

In that case a variable feedback coupling would be necessary to adjust the tube to the conditions for delivering its rated power. An important objection against the latter method lies in the fact that a variable feedback coupling would entail a complicated mechanism, which would very likely also cause microphony and, consequently, noise.

It was therefore finally decided to develop a tube without diaphragms and with a fixed feedback coupling. In spite of the absence of a diaphragm the output cavity can nevertheless be tuned. A simplified cross-section of this tube is shown in fig. 5. The modulating cavity 1 is permanently tuned to the desired resonant frequency. The output cavity 2 is adjusted, while the tube is oscillating, by applying to the collector a large force, denoted by the arrow  $F$ . This force is large enough to strain the cavity wall beyond its elastic limit. When this occurs, the cavity can be deformed to the right amount, at which it remains permanently set. As in the tube shown in fig. 1, the feedback coupling is effected via an opening 10 in the wall between the two cavities.

**Characteristics of the low-noise tube**

The cathode of the tube described here may be either an L cathode<sup>2)</sup> or an impregnated tungsten type<sup>5)</sup>. The life of these cathodes is better than 1000 hours and, compared with tungsten, tungsten-thorium or tantalum cathodes, the heater power they require is low; in the present case it is only 7 W<sup>6)</sup>.

The electrode 6 (fig. 5) with which the electron beam is focused, can be used for amplitude modulation and is therefore generally called the "grid".

As described in the articles quoted in<sup>1)</sup>, a klystron can be made to operate in different modes, i.e. with different electron transit times between the modulating and output gaps. The tube under discussion can operate in three different modes, denoted by A, B and C, which correspond respectively to transit times of  $2\frac{3}{4}$ ,  $2\frac{1}{4}$  and  $1\frac{3}{4}$  periods. The power delivered in these modes is respectively 5, 33 and 200 watts. Each klystron is adjusted to operate in only one of these modes and will not perform properly in other modes.

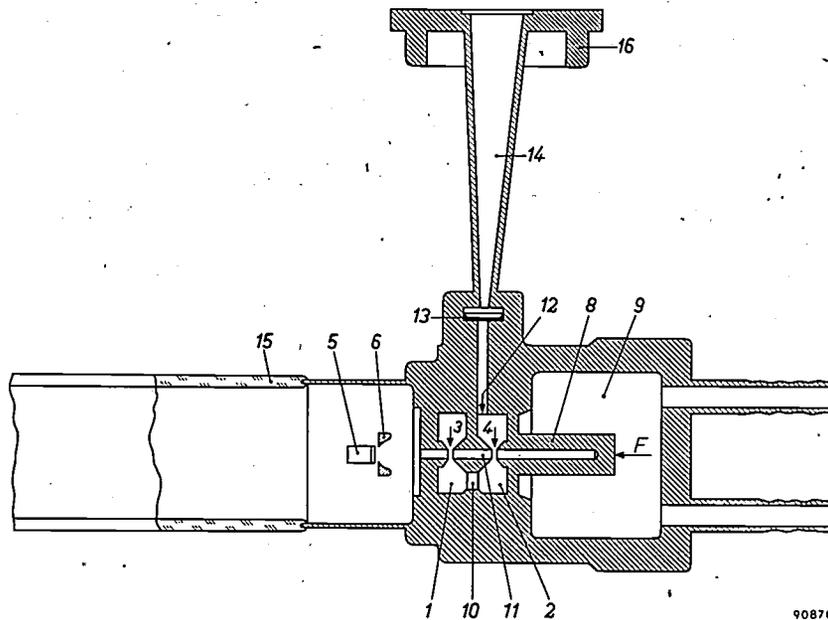


Fig. 5. Cross-section of a low-noise klystron. The figures have the same meaning as in fig. 1. *F* force required to tune the output cavity 2.

By varying the "grid" voltage it is also possible to change the frequency between certain limits. At normal operating conditions this electrode has the same potential as the cathode.

The diode characteristic of the tube, i.e. the collector current plotted as a function of collector voltage, is shown in fig. 6. As for every diode, this characteristic can be represented by the equation:

$$I_c = AV_c^{3/2}$$

In this tube the constant *A* (the perveance) is approximately  $0.25 \times 10^{-6} \text{ A/V}^{3/2}$ . The tube is liquid-cooled and the maximum rate of flow required for water cooling is about 1/2 gallon per minute.

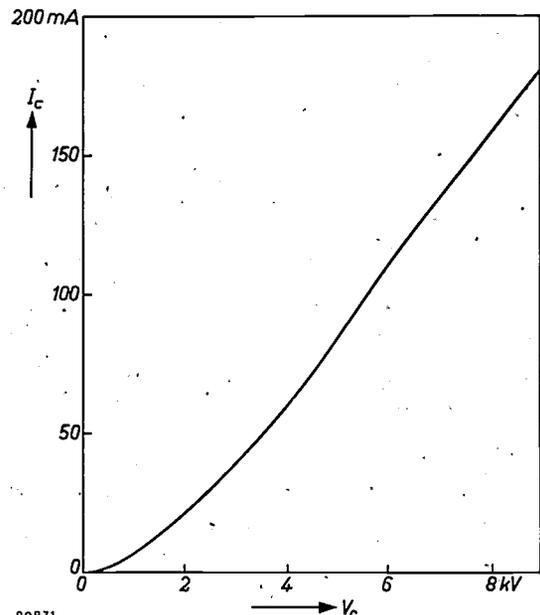


Fig. 6. Diode characteristic of a klystron as shown in fig. 5.  $I_c$  collector current,  $V_c$  collector voltage.

<sup>5)</sup> R. Levi, New dispenser type thermionic cathode, J. appl. Phys: 24, 233, 1953. This cathode will shortly be discussed in full detail in this Review.

<sup>6)</sup> Oxide cathodes, which require an even lower heater power, cannot be used in these tubes because their useful life at the high emission needed would be only a few hours.

Tunable klystrons, as illustrated in fig. 1, do not deliver rated power output over their whole tuning range. This can be seen from the typical power versus frequency characteristic of a tunable tube, at two different modes, shown in fig. 7. The low-noise

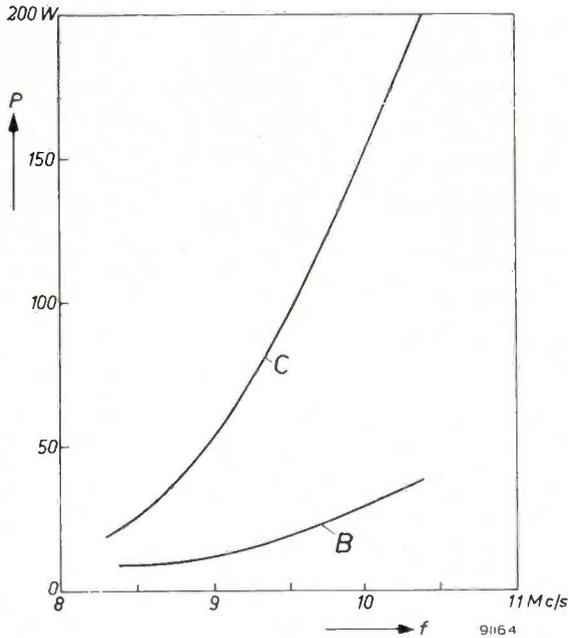


Fig. 7. Power output of a tunable klystron as a function of frequency for modes B and C (electron transit time  $2\frac{1}{4}$  and  $1\frac{3}{4}$  periods respectively).

tube, due to its method of manufacture, gives always its rated power. The frequency where this takes place, can (as indeed the most favourable frequency of a tunable klystron) be fixed at any desired frequency with a maximum error of 10 Mc/s.

The absence of a tuner makes the construction of this tube much simpler than that of tunable types.

This is illustrated in fig. 8, in which are shown an older tunable klystron, a tube with "ruggedized" tuning mechanism and a non-tunable low-noise tube. Because of the remarkable precision with which the new tube can be made for any specified frequency, it is possible to switch over or interchange a series of such tubes for quickly altering a transmitter frequency. This would often be easier than tuning a normal klystron.

The noise of these tubes is so low that we were unable to measure the power spectrum with the system described. It appeared that the waveguides in the measuring system were now more microphonic than the klystrons under investigation.

The elimination of the diaphragms and tuner offers some advantages in addition to the substantial reduction of noise. In the first place it lowers the cost of the tube. Further, it allows more accurate alignment of the electrodes, while the absence of diaphragms leads to a somewhat better heat distribution.

Fig. 9 shows the results of some measurements of the variation of power output and frequency with collector voltage. The curves relate to modes B and C (mode A has been omitted because we are interested mainly in higher power outputs). The power is plotted on a normalized scale, obtained by dividing the power output by the maximum power delivered in each mode. This maximum power is 33 W for mode B and 200 W for mode C, at anode voltages of 4.35 and 8.85 kV respectively. The collector current and the efficiency do not differ much from those of a tunable tube.

The fact that the frequency variation curve reverses its direction is not in accordance with the theory of the electrical phenomena. The assumption that thermal effects in the tube are responsible

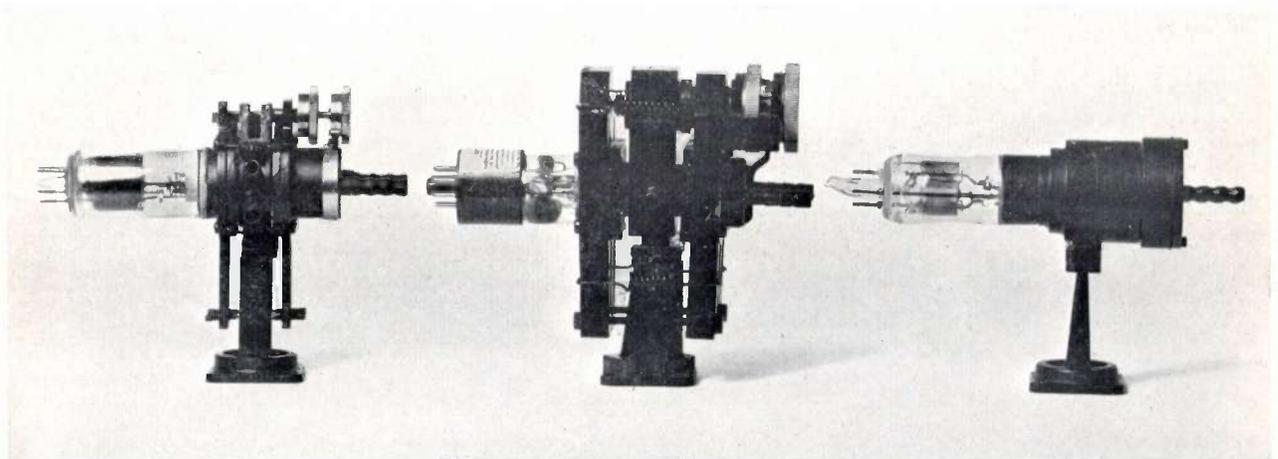


Fig. 8. From left to right: an early type of tunable klystron, a tunable klystron with ruggedized tuning mechanism and a low-noise, non-tunable klystron.

for this phenomenon was made and this was verified by two experiments. Temperature variations influencing the static curves in fig. 9 do not appear if the anode voltage is rapidly varied. For this reason the tube operating in mode *B* was "swept" in anode voltage at a rate of 60 c/s and the frequency

generate frequency-modulation in this way, but this phenomenon can be used for electronic tuning within a certain small frequency range. From fig. 11 it can be seen that the maximum frequency variation is much higher for mode *C* than for mode *B*. The temperature of the coolant has some influence

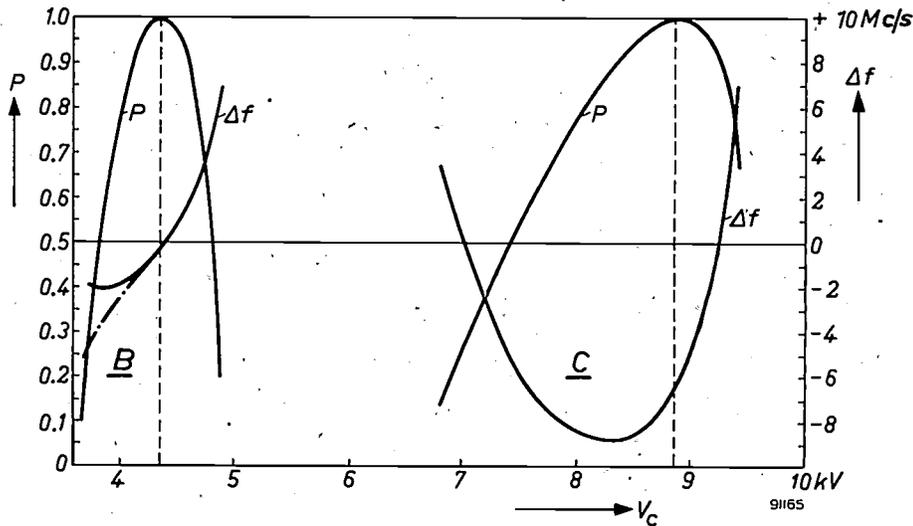


Fig. 9. Power output  $P$  and frequency variation  $\Delta f$  of a low-noise non-tunable klystron as shown in fig. 5, mode *B* on the left, mode *C* on the right. The power is plotted on a normalized scale. The maximum power for *B* is 33 W and for *C* 200 W. The fully-drawn curves are the results of static measurements. The dot-dash curve is the result of measurements in mode *B* with rapidly varying collector voltage.

variations measured through the mode. The result is represented by the dot-dash curve in fig. 9. A reversal of the frequency curve does not occur now.

In a second experiment a small AC voltage with a higher frequency was superimposed on the anode voltage, thus producing frequency modulation. From the spectrum of the output signal the frequency deviation per volt variation in collector potential was calculated. The result is shown in fig. 10, together with the power curve shown already in fig. 9. Whereas it would follow from the static frequency curve that, at a collector potential of about 3.9 kV, small voltage variations would give rise to no variations in frequency, i.e. the modulation sensitivity would be zero, experiment showed that the modulation sensitivity never becomes zero. In fact the frequency variation per volt follows quite closely the slope of the dot-dash curve in fig. 9.

The tube can also be modulated in amplitude by means of the grid voltage. Fig. 11 shows, for modes *B* and *C*, the power in terms of the grid voltage. The latter voltage has also some influence on the frequency. The frequency variation has also been depicted in fig. 11. As this variation is principally due to changes in temperature, it is not possible to

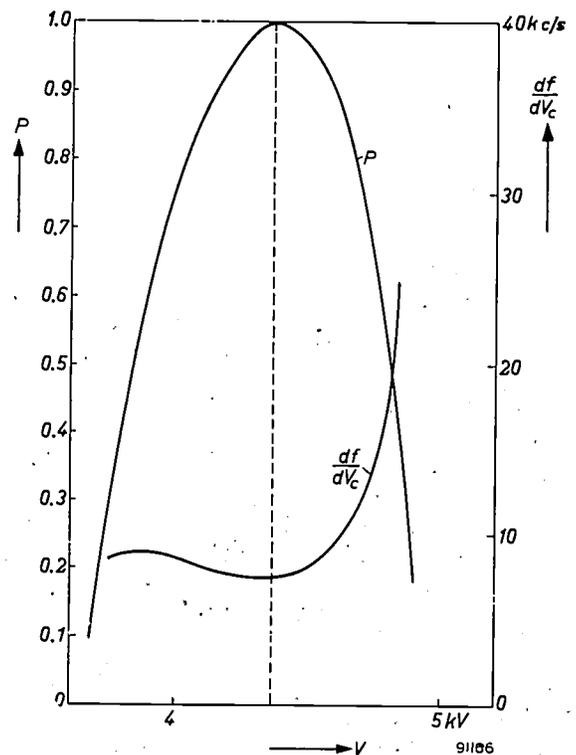


Fig. 10. Frequency variation per V variation in collector voltage,  $df/dV_c$ , as a function of the collector voltage in mode *B*. The power output  $P$  on a normalized scale is also shown; this curve corresponds to that shown on the left in fig. 9.

on the frequency. We found a frequency variation of 0.2 Mc/s per degree variation in temperature. Immediately after switching on, the frequency drift is approximately 2 Mc/s per second for a con-

impedances is given by a Rieke diagram, i.e. a polar diagram in which the modulus and the argument of the reflection coefficient of the load coupled to the waveguide are plotted<sup>7)</sup>. In this diagram lines

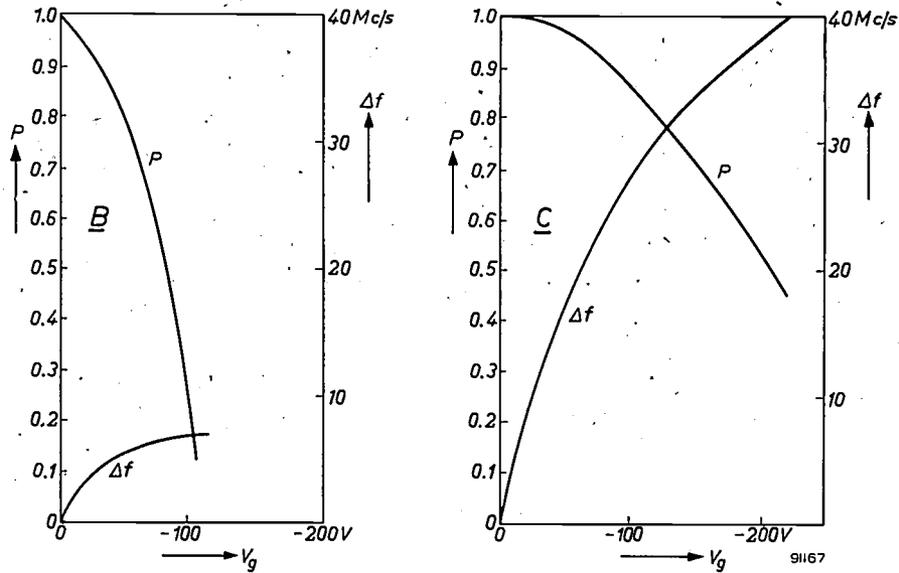


Fig. 11. Power output  $P$  and frequency variation  $\Delta f$  as functions of grid voltage. The power is set out on a normalized scale. The left curve refers to mode  $B$ , maximum power 33 W; the right curve refers to mode  $C$ , maximum power 200 W.

stant coolant temperature. After the warming-up period the effect of ambient temperature upon frequency is extremely slight.

A change in the load impedance affects the output power and also causes a frequency shift. A good impression of the tube's behaviour at different load

of constant power and lines of constant frequency shift are shown. A typical Rieke diagram for a tube operating in mode  $B$  is shown in fig. 12a, and a

<sup>7)</sup> See e.g. D. R. Hamilton, J. K. Knipp and J. B. Horner Kuper, *Klystrons and microwave triodes*, Radiation Laboratory Series, No. 7, McGraw-Hill, New York 1948, Chapter 15.

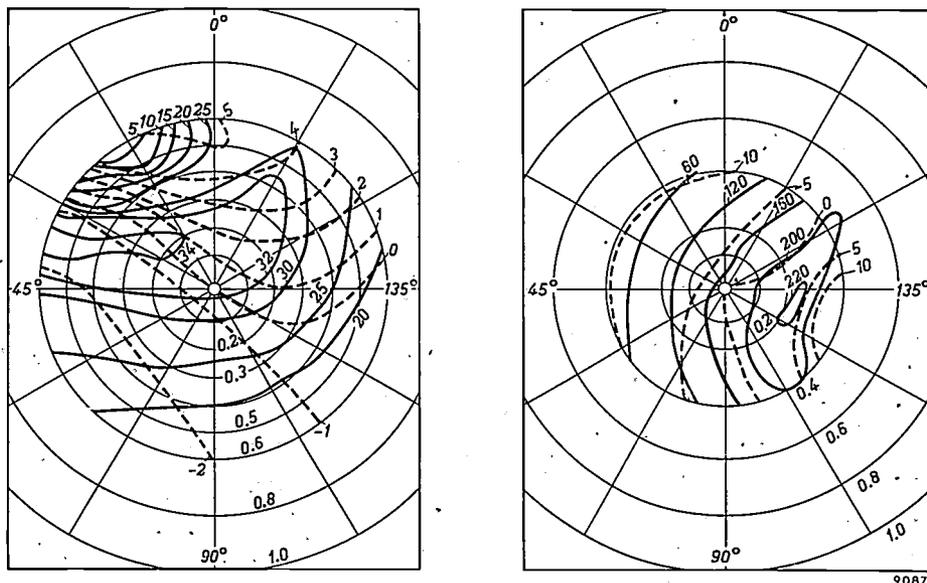


Fig. 12. Rieke diagrams for modes  $B$  (left) and  $C$  (right). The lines of constant power (in W) are fully-drawn; the dashed curves are lines of constant frequency deviation (in Mc/s).

Rieke diagram for a tube operating in mode *C* is shown in fig. 12*b*.

If the load impedance is varied such that the modulus of the reflection coefficient remains 0.2 while the argument passes through all angles between  $0^\circ$  and  $360^\circ$ , a certain frequency variation is produced which is known as the "pulling figure" of the tube. Fig. 12 shows that for modes *B* and *C* the pulling figures are about 4 and 5 Mc/s respectively.

Some tubes, namely magnetrons and some klystrons can be damaged when they are used with a load which is not that which gives the maximum power output. In many cases the window which ensures the vacuum-tight sealing is damaged. In

this respect it is of some interest to mention that the tube described here has worked in mode *B* without difficulties with a load having a voltage-standing-wave ratio of 7. In mode *C* voltage-standing-wave ratios up to 5 have been used.

**Summary.** The flexible diaphragms with which most tunable klystrons are equipped are the most important source of noise in these tubes. Owing to the demand, in a certain application, for a tube with a much lower noise level, a non-tunable klystron has been developed in which these diaphragms are absent. This tube has a very much lower noise level. It can be used in three modes, delivering respectively 5 W, 33 W and 200 W. The tube can be frequency and amplitude modulated by variation of the collector voltage; amplitude modulation can also be effected by the grid voltage. Variation of the grid voltage also permits, via thermal effects, a certain limited tuning of the tube.

## ULTRASONIC MACHINING

### II. OPERATING CONDITIONS AND PERFORMANCE OF ULTRASONIC DRILLS

by E. A. NEPPIRAS\*) and R. D. FOSKETT\*).

534.321.9:621.95

*Machining by means of ultrasonic vibrations is finding widening fields of application. Part I of this article gave an introduction to this technique and described some of the ultrasonic drills developed by the Mullard Research Laboratories. The present (concluding) article considers in some detail the factors which are important with regard to cutting speeds, accuracy and surface finish. A brief comparison is made with other special techniques, notably that of electro-erosion.*

#### Cutting speed

The machining rates attainable with ultrasonic reciprocating tools are affected by many factors, apart from those quantities fixed by the constructional features of the vibrator — oscillatory amplitude, operating frequency and static loading — which have been briefly considered in Part I of this article <sup>1)</sup>. The other factors include: the material of the tool; its shape; its area; the depth of the cut; the physical properties of the work material; and the abrasive properties including the hardness, grain dimensions, the nature of the suspension medium and the concentration of the suspension.

Drilling tests have been carried out under varying operating conditions, in an attempt to assess the effect of all these variables. It was found that with

increase in the static load (*L*), cutting rates (penetration in inches per minute) always increase from zero almost linearly at first to a maximum at some optimum load and then decrease, the curve of cutting rate against *L* becoming asymptotic to the *L* axis (figures 1*a* and *b*). Increase of oscillatory amplitude  $\xi$  at fixed frequency has the effect of increasing cutting speeds and at the same time shifting the optimum load to higher values (see fig. 1*a*). It seems obvious that at least two separate effects contribute to the observed results — a damping effect which becomes important at high values of load, superimposed on a linear law of increase which predominates at low pressures. This becomes more apparent when we study the curves relating to tools of varying areas (fig. 1*b*). The optimum load increases with the tool area, the peak becomes broader, and for tools of very large areas the damping effect becomes unimportant over the range of loads used. For small-area

\*) Mullard Research Laboratories, Salfords, Surrey, England.  
<sup>1)</sup> E. A. Neppiras and R. D. Foskett, Ultrasonic machining, I. Technique and equipment, Philips tech. Rev. 18, 325-334, 1956/57, hereafter referred to as I. Erratum to article I: in footnote <sup>16)</sup>, the subscripts to *k* should read  $\lambda/2$  and not  $y/2$ .

Rieke diagram for a tube operating in mode *C* is shown in fig. 12*b*.

If the load impedance is varied such that the modulus of the reflection coefficient remains 0.2 while the argument passes through all angles between  $0^\circ$  and  $360^\circ$ , a certain frequency variation is produced which is known as the "pulling figure" of the tube. Fig. 12 shows that for modes *B* and *C* the pulling figures are about 4 and 5 Mc/s respectively.

Some tubes, namely magnetrons and some klystrons can be damaged when they are used with a load which is not that which gives the maximum power output. In many cases the window which ensures the vacuum-tight sealing is damaged. In

this respect it is of some interest to mention that the tube described here has worked in mode *B* without difficulties with a load having a voltage-standing-wave ratio of 7. In mode *C* voltage-standing-wave ratios up to 5 have been used.

**Summary.** The flexible diaphragms with which most tunable klystrons are equipped are the most important source of noise in these tubes. Owing to the demand, in a certain application, for a tube with a much lower noise level, a non-tunable klystron has been developed in which these diaphragms are absent. This tube has a very much lower noise level. It can be used in three modes, delivering respectively 5 W, 33 W and 200 W. The tube can be frequency and amplitude modulated by variation of the collector voltage; amplitude modulation can also be effected by the grid voltage. Variation of the grid voltage also permits, via thermal effects, a certain limited tuning of the tube.

## ULTRASONIC MACHINING

### II. OPERATING CONDITIONS AND PERFORMANCE OF ULTRASONIC DRILLS

by E. A. NEPPIRAS\*) and R. D. FOSKETT\*).

534.321.9:621.95

*Machining by means of ultrasonic vibrations is finding widening fields of application. Part I of this article gave an introduction to this technique and described some of the ultrasonic drills developed by the Mullard Research Laboratories. The present (concluding) article considers in some detail the factors which are important with regard to cutting speeds, accuracy and surface finish. A brief comparison is made with other special techniques, notably that of electro-erosion.*

#### Cutting speed

The machining rates attainable with ultrasonic reciprocating tools are affected by many factors, apart from those quantities fixed by the constructional features of the vibrator — oscillatory amplitude, operating frequency and static loading — which have been briefly considered in Part I of this article <sup>1)</sup>. The other factors include: the material of the tool; its shape; its area; the depth of the cut; the physical properties of the work material; and the abrasive properties including the hardness, grain dimensions, the nature of the suspension medium and the concentration of the suspension.

Drilling tests have been carried out under varying operating conditions, in an attempt to assess the effect of all these variables. It was found that with

increase in the static load (*L*), cutting rates (penetration in inches per minute) always increase from zero almost linearly at first to a maximum at some optimum load and then decrease, the curve of cutting rate against *L* becoming asymptotic to the *L* axis (figures 1*a* and *b*). Increase of oscillatory amplitude  $\xi$  at fixed frequency has the effect of increasing cutting speeds and at the same time shifting the optimum load to higher values (see fig. 1*a*). It seems obvious that at least two separate effects contribute to the observed results — a damping effect which becomes important at high values of load, superimposed on a linear law of increase which predominates at low pressures. This becomes more apparent when we study the curves relating to tools of varying areas (fig. 1*b*). The optimum load increases with the tool area, the peak becomes broader, and for tools of very large areas the damping effect becomes unimportant over the range of loads used. For small-area

\*) Mullard Research Laboratories, Salfords, Surrey, England.  
<sup>1)</sup> E. A. Neppiras and R. D. Foskett, Ultrasonic machining, I. Technique and equipment, Philips tech. Rev. 18, 325-334, 1956/57, hereafter referred to as I. Erratum to article I: in footnote <sup>1b)</sup>, the subscripts to *k* should read  $\lambda/2$  and not  $y/2$ .

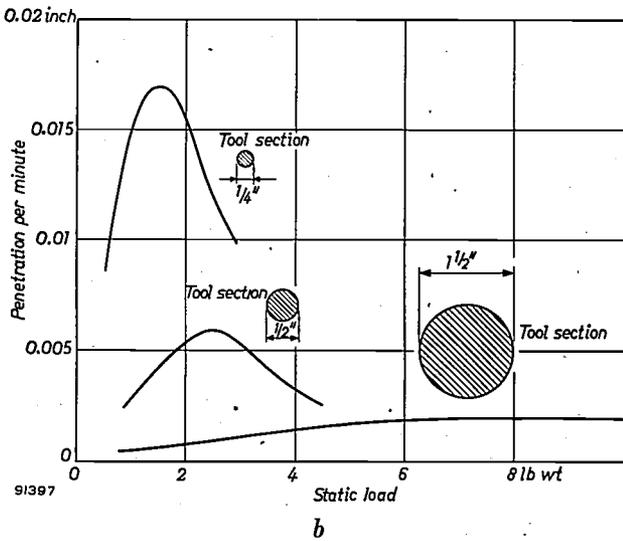
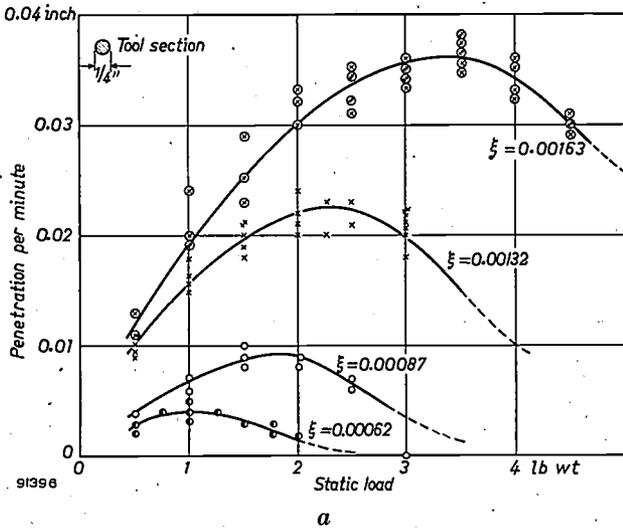


Fig. 1. a) Cutting speed (thousandths of an inch penetration per minute) as a function of static load for four different vibration amplitudes. b) Cutting speed as a function of static load for tool sections of three different areas.

tools, the adjustment of the load for optimum working becomes more critical. Cutting rates are greater for tools of smaller area under otherwise identical conditions, but the increase is not by any means simply related to the cutting area. There is a much closer correlation between the cutting rate and the shape of the tool face — particularly the lateral dimensions when the tool face is long and narrow. This would be expected, the effect being mainly attributable to the difficulty of the abrasive grains penetrating to the centre of a large tool. It also seems that the optimum loading conditions are very dependent on the concentration of the abrasive which reaches the tool face.

The whole position in regard to cutting speeds is complex, but the effects are on the whole what might be expected from the simple mechanics of the process. An increase in the oscillatory amplitude

$\xi$  at constant frequency means a proportionate increase in particle velocity  $\omega\xi$  and in the impulsive forces communicated to the work. The increased amplitude also means that a proportionately greater range of abrasive grain sizes are being brought into use in the grinding process. The cutting speeds would therefore be expected to increase at a rate more than proportional to  $\xi$ , as is indeed the case (fig. 2a). Under conditions of optimum loading, the law is often approximately quadratic. Wide divergencies from this rule occur only when  $L$  does not approximate to the optimum value and in cases where the peak-to-peak oscillatory amplitude of the tool differs considerably from the grain dimensions of the abrasive (see also footnote 8) in I).

The optimum value of external applied load would also be expected to increase with increase in  $\xi$ .

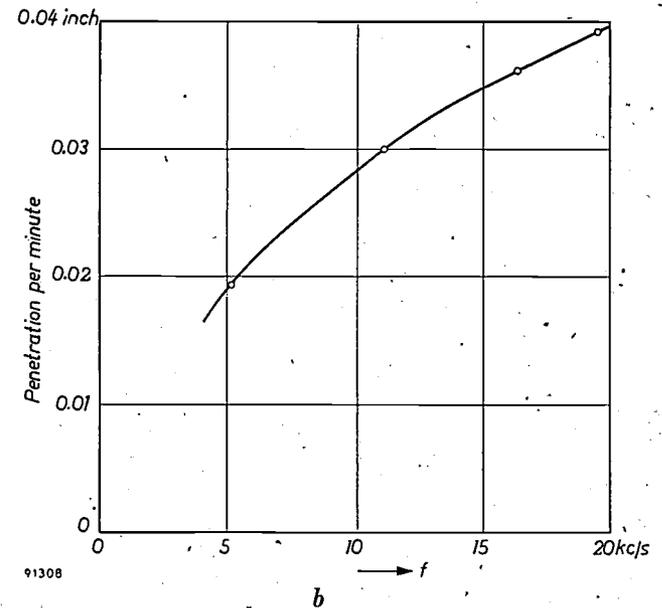
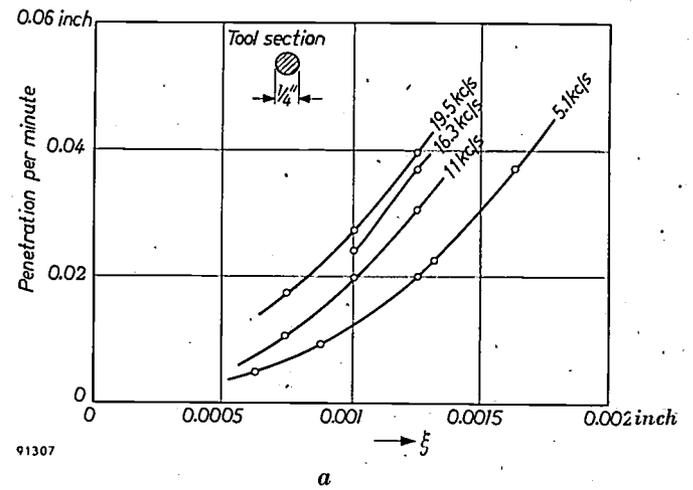


Fig. 2. Cutting rates in glass a) as a function of oscillatory amplitude (peak-to-peak) of the tool face, and b) as a function of the frequency, for a fixed amplitude of 0.00125'' peak-to-peak. These graphs (a) and (b) are reproduced from figs. 1 and 2 of part I.

These predictions take no account of complications which may be introduced due to difficulties of abrasive circulation. Nevertheless, they are well borne out in practice.

In the absence of complicating effects, cutting rates would be expected to increase as the square of the operating frequency at constant amplitude, since the impulse delivered at each impact and the number of blows per second are both proportional to the frequency. Over the range of frequencies 5 to 20 kc/s, however, cutting rates were found to increase slowly with frequency at a rate rather less than linear (fig. 2b), the variation being apparently independent of both oscillatory amplitude and abrasive grit size. This increase is much smaller than would be expected from theoretical considerations. The observations cover the range of frequencies normally used for piezomagnetic transducers. Measurements using other types of vibrator at frequencies around 100 c/s, however, have shown that at these low frequencies the square law relationship between cutting speeds and frequency does hold precisely. The complete curve relating cutting speed with frequency therefore contains a point of inflexion. Beyond this point the overall efficiency drops and there is therefore a theoretical optimum working frequency near this point. However, this would not necessarily be an ideal value to use in practice, since several other practical considerations enter into the choice of operating frequency.

A series of measurements of cutting speeds carried out using tools formed in various common metals of widely different physical properties showed that cutting speeds are not greatly dependent on the tool material; figures for cutting speeds in glass and sintered tungsten carbide, relative to various tool materials are given in Table I.

An advantage is generally gained by using trepanning tools in roughing operations, except where the overall dimensions are small and the hole deep. When tools of very complex section are used, so that the total area of tool surface wall in contact with the liquid is large, considerable viscous damping is imposed on the sides of the tool and this increases with the depth of penetration. For very deep holes the tool should preferably be retracted at regular intervals to achieve an efficient removal of abraded material and help abrasive circulation.

Cutting rates generally increase with increase in grit size of the abrasive powder but only provided the lateral dimensions of the tool are everywhere large compared with the grain dimension. Some typical curves, all referring to the same tool, are shown in fig. 3. There is also a significant correlation between the optimum abrasive grit size and both the oscillatory amplitude of the tool and the static

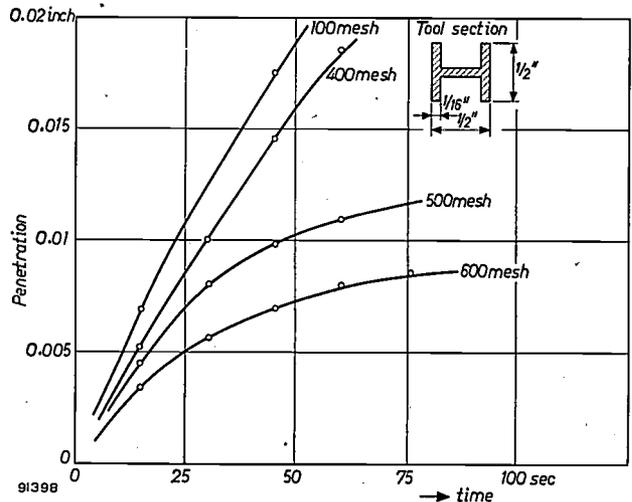


Fig. 3. Typical penetration-time curves showing the effect of abrasive grain size on cutting speed. Grain size is given in terms of British Standard sieves.

Table I. Dependence of cutting speeds on the material of the tool. Cutting rates averaged over about 0.1 inch penetration. (Abrasive: boron carbide; material: soda glass; optimum static load.)

Tool material	Tool shape	Cutting rates, thou'/minute, for			
		$\xi = 0.002''$		$\xi = 0.001''$	
		Abrasive 100 mesh	Abrasive 400 mesh	Abrasive 100 mesh	Abrasive 400 mesh
Copper	Circular, $\frac{1}{2}''$ diam.	63	20	16	6
Brass (BSS 251)	"	90	28	22	8
Mild steel (EN2)	"	86	22	21	7
Silver-steel*)	"	74	22	20	8
Stainless steel (18% Cr, 8% Ni, 0.1% C)	"	71	20	18	6
Tungsten carbide	Triangular, base = $\frac{1}{8}''$	156	100	38	30
Mild steel (EN2)	"	150	102	36	29

\*) C 0.87%, S 0.020%, P 0.025%, Mn 0.28%; tempered in air after oil-quenching from 800 °C.

Table II. Comparative cutting speeds in industrial ceramics (soda glass, using 100 mesh B<sub>4</sub>C abrasive, = 100; tool: mild steel, circular, diam. 1/4").

Work material	Abrasive					
	Boron carbide		Silicon carbide			Alumina
	100 mesh	400 mesh	100 mesh	400 mesh	600 mesh	220 mesh
Chemical porcelain	70	14	53.5	11	2	25
"Temperadex" *)	60	10	45	7.5	—	20
"Faradex" *)	55	7	41	5	—	2
"Vulcanex" *)	190	51	168	44	2	62
Z.Z. porcelain	90	32.5	75	26	2	10
H.T. porcelain	45	24	41	22	—	8
"Frequentite" *)	47.5	27.5	45	26	—	20
Soda glass	100	45	80	37	6	70

\*) Trade names of proprietary materials.

load. It seems to be an advantage to keep the grain dimensions comparable with the peak-to-peak oscillatory amplitude, at any rate when coarse abrasives are used. There is also a very significant variation of cutting speeds with the depth of the cut, as may be seen in fig. 3, and this also involves changes in the optimum static loading conditions and grit sizes.

Measurements of cutting speeds have been carried out in a wide range of materials of widely different physical properties using various abrasives with a wide range of grit sizes. The results, listed in Tables II-IV, show cutting speeds relative to that in soda glass. They show that, in all cases, the drilling

rates depend on the hardness of the abrasive used — cutting rates are always greater with boron carbide than with silicon carbide of the same grit size, and greater with SiC than with alumina. This is rather surprising since hardness is not the only physical property of importance. The shape of the abrasive grains (the number of cutting edges) and their ability to withstand fracture are obviously important factors. It might be thought that in materials which are sufficiently soft relative to these abrasives, the hardness of the abrasive should cease to be important and relative cutting rates would then be determined by the secondary factors

Table III. Comparative cutting speeds in metals. (Tool H-form, 1/2" square, limbs 1/16" thick, see fig. 3; soda glass, using 100 mesh B<sub>4</sub>C abrasive, = 100.) The figures in brackets in Tables III and IV are estimates, interpolated from measurements under somewhat different conditions from those stated above.

Material	Abrasive						
	Boron carbide			Silicon carbide 100 mesh	Alumina 220 mesh	Sand (grit size 0.012")	Diamond powder (grit size 0.001")
	100 mesh	220 mesh	400 mesh				
Soda glass	100	90	77	85	65	47	90
Brass (common yellow)	6.6	5.6					
Die steels							
(i) K.E. 672 (R. 66)	1.4	1.3					
(ii) C.S.K. (R. 62)	3.9	3.6					
(iii) K.E. 672 (R. 61)	2.2	2.1		1.48	0.1	0.1	
(iv) K.E. 672 (R. 58)		1.7					
Stainless steels							
(i) 18% Cr, 8% Ni, 0.1% C	2.1	1.9					
(ii) 3.5% Cr, 8.4% W, 0.35% V, 0.3% C	1.2	1.1					
Carbon-chrome bearing steel (heat-treated)	1.4						
Sintered tungsten carbide (R. 76)	4.1	3.5		2.55	0.2	0.2	4.3
Tungsten	4.8	4.3					
"Stellite"	4.0	3.7					
Germanium single crystal	(31)			(28)			
Titanium	(4.0)						
Beryllium	(7)						

Table IV. Comparative cutting speeds in various brittle materials. (Tool H-form,  $\frac{1}{2}$ " square, limbs  $\frac{1}{16}$ " thick, see fig. 3; soda glass, using 100 mesh B<sub>4</sub>C abrasive, = 100.)

Material	Abrasive			
	Boron carbide			Silicon carbide 100 mesh
	100 mesh	200 mesh	400 mesh	
Soda glass	100	90	77	85
"Hysil"	73	66	54	
B9 borosilicate glass	86			
Ferroxcube IIC	37			34
Ferrodure (demagnetized)	(32)			
Quartz crystal	(57)			
Fused alumina	19			
Synthetic sapphire	19			
Synthetic ruby	18			
Flint stone	(72)			
Barium-titanate ceramic	110			109
Ceramic 507	(38)			35
Carnet	(58)			
Felspar	(40)			
Spinel	(48)			
Slate	67			
"Mycalex"	(240)			(200)

mentioned above. If this was the case, silicon carbide would provide the fastest rates. If hardness is really the most important physical property of the abrasive determining cutting rates it should be possible to use an ultrasonic machining test to grade abrasives in a reliable scale of hardnesses. A soft brittle material would be used as a standard sample test piece. Taking soda glass as a very brittle and comparatively soft material, the figures given in the tables rank boron carbide, silicon carbide and alumina in the order 100:80:72. The Knoop hardness figures for these abrasives are 2250, 2100 and 1650 (100:93:73).

From a study of the data in Tables II, III and IV, it is clear that brittleness is the most important single property of the workpiece which determines its machinability. (For the present purpose brittleness is taken to mean the ability to chip pieces up to 0.001 inch in size from the workpiece.) It should be noticed that many materials which are ductile and malleable at ordinary temperatures, and therefore not easily machined ultrasonically, become brittle when the temperature is reduced sufficiently, so that the technique in fact possesses a considerably wider potential field of application than might be thought from the measurements recorded here which, of course, all refer to normal room temperature.

It will be seen that cutting rates are greatest for

glasses and certain ceramics. In metals, with the exception of germanium, speeds are lower, generally by a factor of some 25-100. However, even in tough hard metals not noted for their brittleness, cutting speeds are generally high enough to make the ultrasonic technique industrially interesting.

Experience has shown that the most convenient method of feeding the abrasive to the tool in practical drilling operations, particularly when long jobs are involved, is to operate a continuous circulation system, employing a considerable volume of slurry so that only infrequent bulk changes of the abrasive are necessary. When the abrasive is fed externally to the tool it is pumped on to the work from one or more jets; for large areas it is essential to use several jets to ensure an even circulation of abrasive in the cut. In deep holes a fast continuous jet feed generally produces rather faster cutting rates as compared with manual feed, since the jet aids the abrasive circulation. An alternative method of applying the abrasive to the work, which overcomes the difficulties encountered in deep drilling operations, is to force-feed the abrasive suspension through the tool itself. This has been found a very valuable technique but it is of course only applicable to trepanning operations.

For efficient operation, the usable concentration of the abrasive is limited by the ability of the mixture to flow easily under the tool and by the danger of a very concentrated suspension drying out. The optimum concentration will vary with the working conditions, abrasive grain size and suspension medium but as a general guide it should form a paste of syrupy consistency. Typical curves showing the variation of cutting speed with abrasive concentration are given in fig. 4.

In deciding the best type of liquid to use as the carrying medium for the abrasive, the indications are that a low viscosity is a desirable property, particularly in piercing operations, since the abrasive suspension must flow in a restricted path and the intensity of cavitation (on which depends the efficiency of abrasive circulation) is also a function of the viscosity of the medium. The liquid should also have good wetting properties with respect to the work material, the tool, and the abrasive; it should presumably also have a high density, approaching that of the abrasive (in order to retain it in suspension) and, for efficient cooling, a high thermal conductivity and specific heat. Experimental work carried out using a wide range of available liquids as carrying media for the abrasive has largely verified these ideas. Water, with or without small additions of surface activating agents,

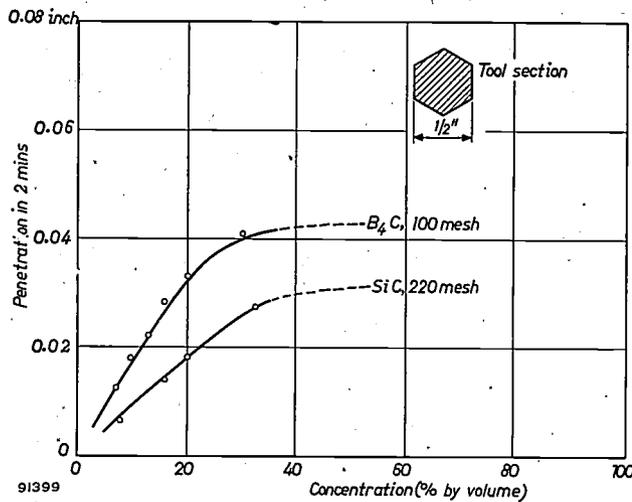


Fig. 4. Cutting speed in glass, as a function of abrasive suspension concentration, for two different abrasives.

is about the best liquid available and far more satisfactory than most oils, even thin oils.

Measurements have also been made of the wear produced in the tool (see Table V) under similar conditions to those noted in Table I. Of the materials tested the relative wear is least in the case of silver

steel. It seems that, for minimum wear, a hard tough steel is desirable. Whether, in any particular case, the advantage to be gained in using a material of this sort rather than a softer steel is likely to outweigh the disadvantage resulting from the greater difficulty in fabricating the tool is another question. Tools of copper and similar very soft metals are definitely not suitable for piercing work, not because wear is excessive or that cutting speeds are slow, but because plastic deformation of the tool at high vibration intensities causes distortion of the tool face (mushrooming).

To conclude this section, the following specific example may be quoted for the cutting rate obtained in machining a typical shaped hole in a brittle material: Using a spline trepanning tool of mild steel, about 1/2 inch diameter, similar to that shown in the photograph of fig. 11 in I, a hole was cut through 1/4 inch thick sintered tungsten carbide in about 42 minutes. In this operation the peak-to-peak oscillatory amplitude of the tool was 0.0025 inch, 220 mesh boron carbide powder in water was used as abrasive and the static loading was adjusted to about the optimum value, 3 1/2 lbs. wt.

Table V. Tool wear. (Abrasive: boron carbide 100 mesh;  $\xi = 0.002''$ ; optimum static load.)

Tool material	Tool shape	Work material					
		Glass			Tungsten carbide		
		Longitudinal tool wear (inches)	Total penetration in workpiece (inches)	Tool wear as % of stock removal	Longitudinal tool wear (inches)	Total penetration in workpiece (inches)	Tool wear as % of stock removal
(a)							
Copper	Circular, 1/2" diam.	0.0025	0.520	0.48	—	—	—
Mild steel (EN2)	"	0.018	1.850	1.0	0.110	0.125	88
Silver steel *)	"	0.0025	0.546	0.46	0.012	0.046	26
Stainless steel (18% Cr, 8% Ni, 0.1% C)	"	0.008	1.150	0.7	0.016	0.045	35
Brass (BSS 251)	"	0.021	1.250	1.68	0.175	0.125	140
Sintered tungsten carbide	Triangular, base = 1/8"	0.0015	1.510	0.1	0.138	0.125	110
(b)							
Mild steel	Circular, hollow, int. diam. = 1/8", ext. diam. = 1/4"				0.156	0.125	125
Mild steel	H-form, 1/2" square, limbs 1/16" thick	0.053	0.024	220			
Mild steel	Circular, hollow, int. diam. = 0.33", ext. diam. = 0.39"				0.170	0.075	222
Brass	Extrusion shape T-form, height of T = 16 mm				0.256 (average)	0.125	205

\*) C 0.87%, S 0.020%, P 0.025%, Mn 0.28%; tempered in air after oil-quenching from 800 °C.

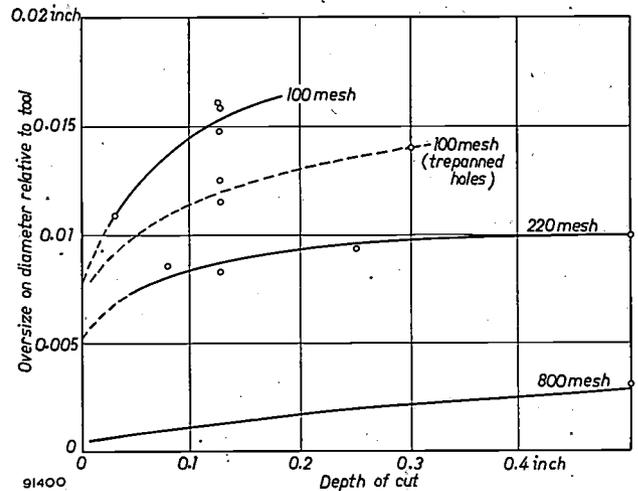
The work on cutting speeds reported above refers to piercing operations. In lapping and re-passing work in which the wearing is produced essentially by frictional action, the position is much simpler. Difficulties due to abrasive circulation and complications concerning the relation between abrasive grit dimensions and tool size, oscillatory amplitude and penetration, do not arise. Also, experience shows that in this type of operation there is no optimum static load related to the other variables. Provided that the cutting is done essentially by frictional forces, cutting rates increase continually with the applied load, until a point is reached when the tool tends to stick in the hole and the flow of abrasive is impeded. In all operations of this sort, it is a great advantage to retract the tool at intervals to allow a better flow of abrasive between tool and work. If this is not done, the maximum usable load may be very small in deep holes, unless the oscillatory amplitude is much greater than the abrasive grain dimensions.

An important advantage of the high frequency machining technique is that under normal operating conditions it leaves no residual mechanical stresses in the material which is being cut. This is shown by the fact that even in the most brittle easily-fractured materials, the pieces chipped off in the machining process are small compared with the abrasive grit dimensions, showing that the stresses communicated to the workpiece are extremely localized even though the particle acceleration of the tool is generally very great. (The acceleration  $\omega^2\xi$  is, for example, about 25 000 g for a frequency of 16 kc/s and a peak-to-peak amplitude of  $\sim 0.002''$ .)

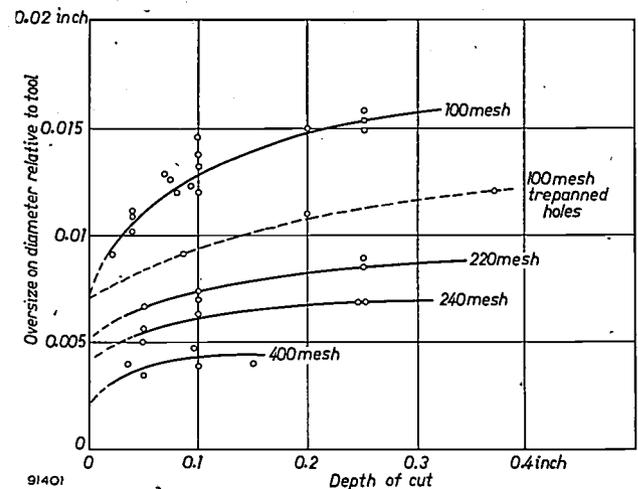
### Accuracy

Since the abrasive grains must necessarily pass down the sides to get under the tool, the resulting hole is oversized relative to the tool, the oversize being greater at the entry and getting progressively less towards the exit for through holes. When close dimensional tolerances are required it must be possible to predict the oversize quantitatively.

The variations of oversize with the depth of the cut and grit size of abrasive, under approximately optimum load conditions, have been measured and are shown in *figs. 5a* and *b*. The oversize on radius is initially always a little greater than the lower limit of the grit dimensions of the abrasive. The oversize at the entry increases rapidly, as the hole becomes deeper, to a value which corresponds fairly well to the upper limit of the grit dimensions. Thereafter the increase is comparatively slow, being a function of both penetration and time, and relatively



a



b

Fig. 5. Oversize on diameter as a function of depth of cut, for various abrasive grain sizes, *a*) in tungsten carbide, *b*) in soda glass. Abrasive in both cases was boron carbide.

greater for the smaller grit sizes. The oversize is seen to be a little greater in tungsten carbide than in glass for the same depth of cut, showing that the ratio of lateral to longitudinal wear is greater in the former case.

Measurements show that drilled *blind* holes are always tapered towards the bottom in such a way that the oversize at the bottom is only of the order of the smallest of the abrasive grains present in the suspension. This means that probably few of the larger grains ever reach the tool face in such a drilling operation.

Apart from the abrasive, other factors contributing to the oversize obtained are those relating to the precision of the mechanical arrangements: 1) the accuracy with which the transformers and tools are aligned with respect to the transducer, 2) the accuracy with which the slide mechanism allows motion

of the drill in a vertical direction only, and 3) the relative perpendicular alignment of the drill and the worktable. Deficiencies in 1) and 3) will result in the tool entering the work in a slanting direction so that the hole drilled out by a circular tool will show ovality of an amount depending on the depth of penetration.

Since cutting rates are so much greater when coarse abrasives are used, while good surface finish and close dimensional tolerances both demand fine grit size, the drilling operation must be performed in several stages when very accurately dimensioned holes are required. A hole is first drilled through using an undersized tool of dimensions which can be estimated from a knowledge of the depth of the cut and grain size of the abrasive to be used. Then a second repassing or sizing operation is carried out, for which the tool will be cut only slightly undersize, a fine abrasive being used, the precise grit size depending on the surface finish required. The repassing process will be rapid since if the roughing operation has been correctly carried out it will be necessary to remove only a very small volume of material. When great accuracy is required, or very deep holes to a reasonable tolerance, the hole can be drilled in three operations, the final one being done with a close tolerance tool and a polishing grade of abrasive, such as 2000 mesh.

The order of dimensional accuracy obtainable in single-stage drilling operations is such that, even when using coarse (220 mesh) abrasive powder, the dimensions of the cut can be predicted to within

dimensional accuracy achieved in glass and tungsten carbide, using 800 mesh repassing powder, can be better than  $\pm 0.0005$  inch.

To illustrate the accuracy obtainable by ultrasonic machining in the important field of die making, two typical examples will be described. The first of these is a gauge in tungsten carbide designed to check splined shafts. The shaft cross section and that of the gauge are shown in *figs 6a* and *b*; the gauge is also illustrated in the photograph of *fig. 6c*. The drilling was carried out as a 3-stage operation. A circular hole was first cut in the blank, using a trepanning tool and a roughing grade of boron carbide. The diameter of the hole obtained by this operation was approximately 0.015 inch less than the required 0.530 inch root diameter of the finished gauge. A solid splined tool having the shape of the required finished hole but 0.005 inches undersize was next used in a second roughing operation. The resulting hole was approximately 0.0015 inches radially undersize relative to the required dimensions. The work was now considered ready for finishing and a final tool 0.0005 inches undersize was used together with a fine grade of boron carbide.

The second example to be described, a blanking die for valve anodes, is illustrated in *figs. 7a* and *b*. The die was cut from a  $\frac{1}{8}$  inch (3 mm) thick tungsten carbide plate. In this case, one roughing cut was followed by two finishing operations. The tools were made successively 0.010, 0.004 and 0.001 inch undersize on all internal linear dimensions and were brazed on to cylindrical brass half wave

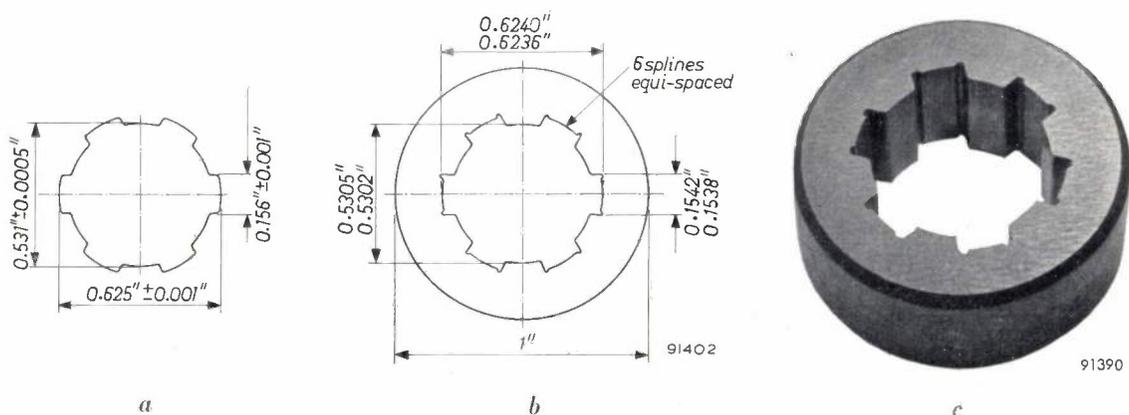


Fig. 6. a) Cross-sections of splined shaft. b) Dimensions of tungsten carbide ("no-go") gauge. c) Photograph of finished gauge.

about  $\pm 0.002$  inch if a tool is used which is considerably longer than the hole depth. (The tool will then follow through after penetration and remove, by lapping, inaccuracies due to tool wear.) For many applications this degree of precision may be sufficient. For two-stage drilling operations, the

stubs. In order to make use of the slight taper obtained in a normal drilling operation, the work was so arranged that the face clamped to the work table should become the upper face of the die. By a final lapping operation using the largest tool and 800 mesh abrasive, the die was brought very

accurately to size so that it was just possible to press the punch through by hand. The resultant clearance between punch and die was very small — of the order of 0.0003 inch. Sections of the die wall taken after each ultrasonic operation are shown in fig. 7c. These show clearly the dimensional accuracy obtained in the operation.

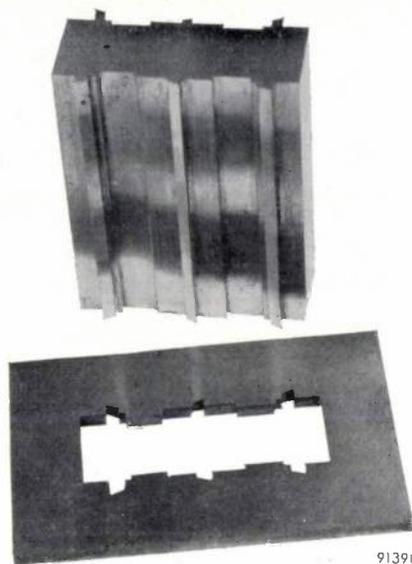
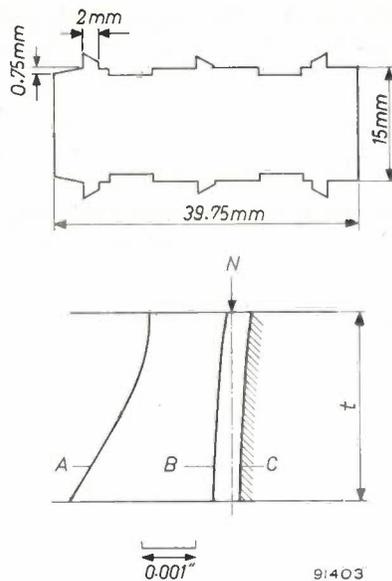


Fig. 7. a) and b) Photograph and sketch of tungsten-carbide blanking die for valve anodes. c) Contours of the die wall, *A* after roughing out, *B* after second operation and *C* after the final ultrasonic machining operation. *N* shows the specified nominal size. The thickness *t* of the die is drawn on a much smaller scale than that used for the horizontal distances.

To summarize we may say that the order of dimensional accuracy depends only on the two factors mentioned above — the abrasive grit size and the mechanical precision inherent in the mounting assembly — and is not restricted by any basic limitation imposed by the technique itself. With regard to the precision of the mounting, the two main problems are 1) ensuring that the drilling tool is fixed with its axis parallel to the line of movement of the drill head or the work; 2) aligning a tool for a finishing operation with the cut already produced by a roughing tool. Both these questions, which are closely related, are being investigated.

Even for cutting circular dies, the vibration technique possesses a considerable advantage in speed over conventional methods. In such cases, both the cutting speeds and circularity of the die hole can be improved by rotating the die at a slow constant rate during the drilling. The essentials of a rotatable die holder used for the production of tungsten carbide wire-drawing dies are shown in

the sketch of fig. 8. The die blank is supported centrally in a circular metal container carrying the abrasive suspension. The die holder carries holes at the base to allow the continuous circulation of abrasive through the die hole. The tool is supported on the die with the appropriate static load and the container rotated at a slow rate (about once every



5 seconds). The abrasive is circulated either manually or by means of a fixed vane dipping into the slurry.

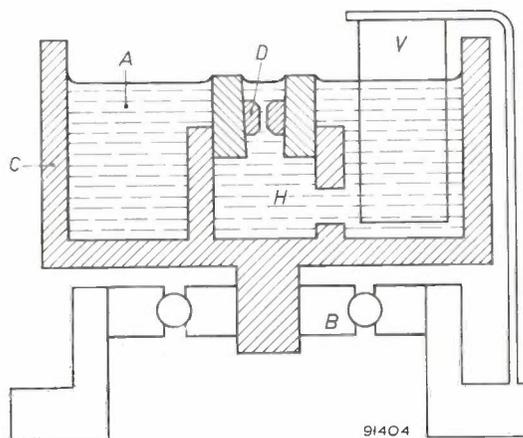


Fig. 8. Mechanical set-up for the ultrasonic drilling of tungsten carbide wire-drawing dies. *A* abrasive suspension, *B* bearing, *C* rotatable metal container, *D* die blank, *V* sloping vane to aid stirring of abrasive, *H* holes in die holder to provide path for abrasive circulation.

### Surface finish

Measurements have been made of the quality of the surfaced finish obtained on the bottoms and sides of roughed and repassed holes in glass and tungsten carbide. These measurements were carried out with the aid of a Talysurf roughness tester which plots a trace of the surface irregularities and records a centre-line average figure for the roughness height over the final 0.3 inch of surface examined. The following general conclusions can be drawn from all the measurements recorded.

The quality of the surface finish is closely related to the abrasive grit size as shown clearly in the four traces of *fig. 9* referring to 100, 400, 600 and 800 mesh boron carbide powders, the measurements being taken at the bottom of blind holes drilled out under approximately optimum drilling conditions. For the coarse abrasives the peak-to-peak irregularities sometimes amount to as much as 0.001 inch; for the fine powders they are never more than about 0.0001 inch. Trace (*d*) is interesting in that it shows what appears to be a small cavitation channelling mark. These are rarely found at the bottom of blind holes.

with abrasive flow. Increasing the load results in a reduction in the size of abrasive grains which can reach the tool face. Consequently, when coarse abrasives are used the roughness height on the sides of blind holes is greater than at the bottoms.

The roughness height does depend somewhat on the oscillatory amplitude of the tool: it increases slowly with amplitude under optimum drilling conditions.

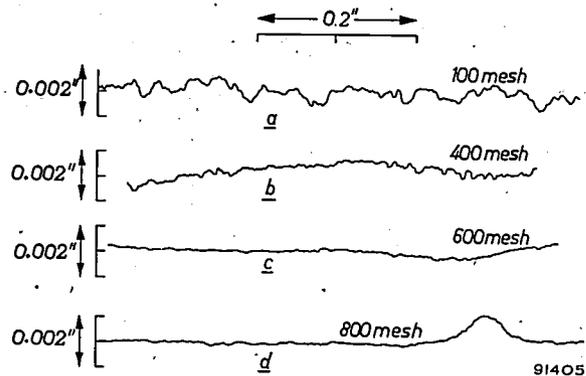


Fig. 9. Surface traces over the bottoms of blind holes in glass, bored with boron-carbide abrasives. (Talysurf roughness tester. Note difference in vertical and horizontal scales.)

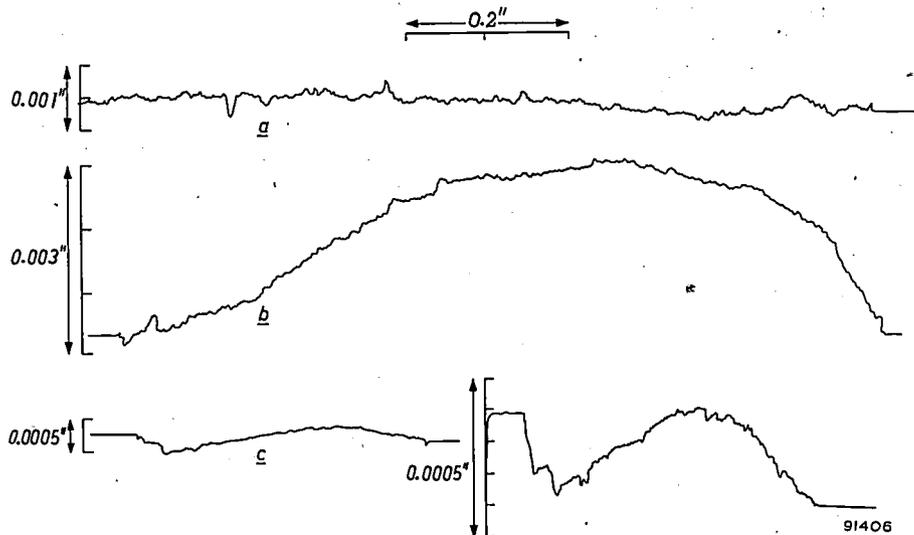


Fig. 10. Surface traces for blind holes in tungsten carbide, bored with boron-carbide abrasives.

a) Bottom of surface cut, tool penetration 0.005"; 100 mesh abrasive.

b) Bottom of hole 0.1" deep, 100 mesh abrasive.

c) Side of repassed hole, 800 mesh abrasive. The trace on the right refers to the same region as that on the left, but is on a 5× enlarged scale.

The quality of the surface finish appears to be quite independent of the static loading conditions. This is a rather surprising result; it might be expected that the roughness height should increase approximately linearly with the applied load in the same way as cutting speed. The explanation appears to be that there is a compensating effect associated

The traces of *fig. 10* refer to cuts in tungsten carbide. Traces (*a*) and (*b*) refer to the bottom of blind holes, using 100 mesh boron carbide abrasive (amplitude = 0.002 inch, under approximately optimum operating conditions). (*a*) is for a surface cut, the tool having penetrated only about 0.005 inch. The surface irregularities are greater than in

(b), a similar trace for a hole 0.1 inch deep. This is an indication that when the abrasive circulation depends on the motion of cavitation streamers (viz. when hole depth  $\gg$  clearance between tool and work, i.e.  $\gg$  abrasive grain size), the largest grains never reach the tool face. In (b) the large-scale modulation of the trace represents a surface undulation caused by irregular tool wear. Trace (c) refers to the side of a repassed hole in a tungsten carbide slab, using 800 mesh boron carbide powder.

Fig. 11 shows a plot of the measured surface roughness height as a function of the abrasive grit size in glass and tungsten carbide. In piercing operations the roughness height at the sides is of the same order as that obtained at the bottom except when the coarsest abrasives are used. In repassing operations, for all practical purposes the surface finish can be taken to be dependent only on the grit size of the abrasive and independent of the oscillatory amplitude and static load. Under the same operating conditions, the roughness heights in glass and tungsten carbide (bottoms of holes) are approximately in the ratio 7:1, although cutting speeds are about in the ratio 24:1. The surface traces show clearly that stock removal occurs essentially by a chipping action in both glass and carbide.

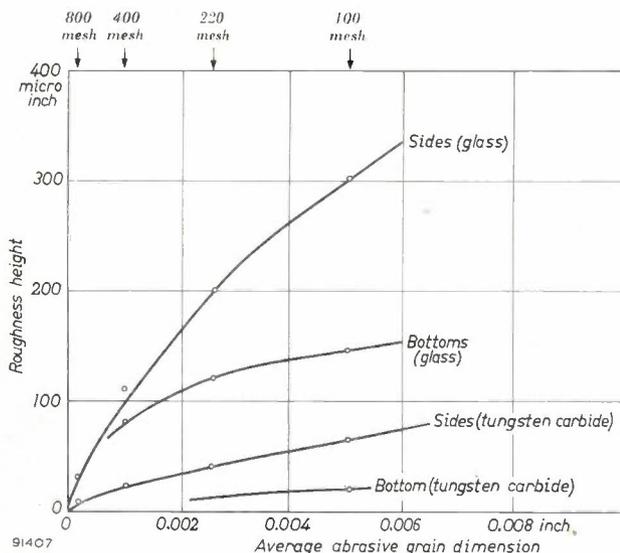


Fig. 11. Surface roughness height as a function of abrasive size for holes in glass and tungsten carbide.

The quality of the surface finish obtained on the sides of holes is sometimes by no means uniform. The rougher patches consist of channels cut into the surface of the material by the action of cavitation streamers drawing abrasive grains between the tool and the work. These markings have already been referred to in Part I of this article (see also

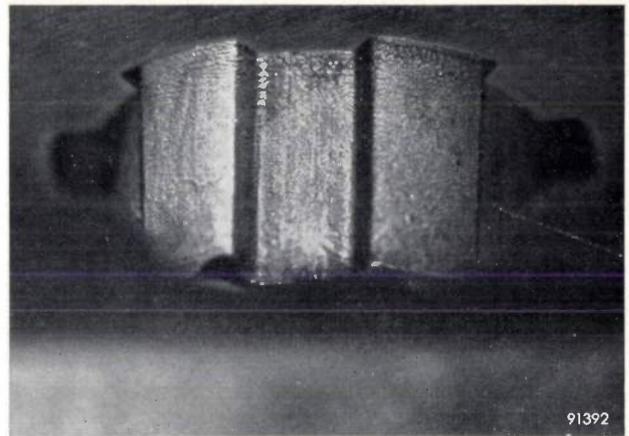


Fig. 12. Micrograph showing cavitation markings on the surface of a cut in tungsten carbide.

fig. 9d above). Their nature is shown clearly in the micrograph of fig. 12. The width and depth of the channels seem to be related to both the grain size of the abrasive and the width of the streamers themselves. With coarse abrasive the depth of the channels is about the order of the grain dimensions, and although they become shallower for finer abrasives they are rarely less than a few thousandths of an inch in width and depth. As already stated in I these markings can be prevented from forming by carrying out the machining operation sufficiently rapidly: the cavitation streamers are then unable to form a stable pattern and wear due to this cause is thereby distributed randomly over the surface.

#### Comparison with other techniques

Tools based on ultrasonic vibrators have rendered possible new types of machining operations, jobs which had not previously been attempted on account of their extreme difficulty or tediousness. In some of its potential uses, however, the technique is in competition with other methods. This is true of most applications relating to metals, including the fruitful field of die-making. Apart from sintering followed by hand-lapping techniques, there are other more efficient modern processes now being applied to this problem of the accurate machining of complex shapes in hardened steels and carbide. These include the various spark and electro-arc techniques.

Briefly, the ultrasonic machining technique is characterised by rapid cutting rates combined with the comparative ease with which accuracy and extremely good surface finishes can be achieved. No heat is generated in the work and consequently no changes are introduced in the physical structure of the material. The equipment is safe to handle and requires skilled labour only in making and

setting of the actual tools; these are, however, cheaply made in a comparatively soft metal. Cheap abrasives are used and expensive diamond bort is needed only for drilling diamond. It is also important to notice that the technique is applicable to all brittle materials and is not limited by any other physical property of the substance. Extreme hardness of the work material does not constitute a limitation to its machinability, but it is, of course, necessary to choose an abrasive which is as hard or preferably appreciably harder.

For forming irregular-shaped holes and depressions, patterning and multiple cutting in brittle materials, we can say that applied to non-conducting substances the vibration technique virtually has no competitor. When applied to conducting materials on the other hand the ultrasonic method is in competition with the new electro-erosion processes. The chief competitor to vibration machining techniques is the "electrosparking" process<sup>2)</sup> developed commercially in a number of countries, including America, Britain, Switzerland, Czechoslovakia and Russia. Fundamentally the process depends on producing very high localised instantaneous charge densities at the surface of the workpiece by initiating a succession of spark discharges across the space between the tool and workpiece, this being filled with a non-conducting liquid. In this process the spark erosion produces saucer-shaped depressions in the surface of the workpiece which are larger the greater the energy of the spark, so that the quality of the surface finish is in inverse ratio to the cutting speed. Although the method is capable of extremely high cutting speeds, greater than at present can be achieved by the vibration technique, it is not possible to rapidly achieve a good quality finish. It seems that to obtain even a

reasonably good finish it is necessary to perform the machining in several successive operations, of which the final one may be an extremely lengthy process. The ultrasonic method is not limited in this way, a reasonably good finish being obtained even in roughing operations. For this reason it has been suggested that it may be possible to combine the two techniques, using the spark machine for roughing where any considerable bulk of material is to be removed and the ultrasonic abrasive process for finishing. Some experiments have been carried out in this direction with encouraging results.

Both these new machining techniques are still very much in the development stages and it is too early to attempt a rigid assessment of their comparative value. It seems that to a great extent the techniques are complementary rather than competitive even when applied to conducting materials, while for non-conductors the ultrasonic method has no serious competitor.

---

**Summary.** This second article on ultrasonic machining deals in some detail with cutting speeds, accuracy and surface finish. The various factors affecting cutting speeds are considered with the help of graphs and tables. Cutting speeds increase almost linearly with static load until an optimum value (dependent on vibration amplitude and frequency) is reached. For tools of small area the adjustment to optimum load is critical. Within the range of usable amplitudes and at ultrasonic frequencies, cutting rates increase rather more than proportionally with amplitude and rather less than proportionally with frequency. Cutting speeds are higher with coarser abrasives up to a grain size limited by the vibration amplitude. Considerable data is given of actual cutting speeds in various metals and brittle materials (glass, ceramics, minerals, etc) using different abrasives of various grain sizes. Cutting speeds vary from a few thousandths of an inch to some tenths of an inch per minute according to the work material, abrasive, tool material, etc. Tool wear is also discussed briefly. The effect on accuracy of abrasive grain size and mechanical precision of the drill mounting are discussed. Two examples are given illustrating the accuracy obtainable (of the order of  $\pm 0.0005''$ ). Surface finish and its relation to abrasive grain size and vibration amplitude is also discussed. Ultrasonic machining has no competitor where hard non-conducting materials are concerned; for conducting materials it is less rapid than some electro-erosion techniques but is capable of higher accuracy and smoother finish.

<sup>2)</sup> See for example I. Koncz, Electric spark machining, *Gépész (Hungary)* 1, 388-411, Sept.-Oct., 1949 (abbreviated translation in *Engrs. Digest (London)* 11, 108-111 and 165-168, 1950); also D. W. Rudorff and H. Drubba, *Research (London)* 7, 216-220, 1954.

## CATHODE-RAY DISPLAY OF COMPLEX QUANTITIES AT VARYING FREQUENCIES

621.317.34:621.317.755

It is often required to know precisely how a complex physical quantity behaves throughout a certain range of frequencies. Examples of such quantities are:

- 1) The complex ratio of input voltage or current to output voltage or current in electrical four-poles such as amplifiers, band pass filters and transistors, the ratio being expressed by a polar phase-amplitude diagram or Nyquist curve <sup>1)</sup>.
- 2) The complex ratio of input voltage to input current, or *vice versa*, at a terminal pair of an electrical or electro-mechanical two-pole or multi-pole.

As examples of electro-mechanical two-poles, mention may be made of electro-acoustic transducers such as piezo-electric and piezomagnetic vibrators <sup>2)</sup>; it is mainly in the vicinity of their mechanical resonances that the behaviour of these devices is of interest.

To make rapid but nevertheless reasonably accurate determinations of complex quantities at various frequencies, we have developed an apparatus which automatically gives a picture of the complex vector diagram, for an adjustable frequency range, on the screen of a cathode ray tube <sup>3)</sup>, this being in fact a television picture tube. Until now a bridge method has had to be used for determining the complex impedance or admittance diagram of an ultrasonic crystal vibrator — to give one example — and this method means determining the diagram point by point. The procedure takes many hours, during which various quantities have to be kept strictly constant. The new method allows the diagram to be made visible in a few seconds, and it indicates immediately how the diagram changes when something is modified in the electrical network or the acoustic load of the test object.

The new method is partly based on the principle described in the article named in footnote <sup>1)</sup>. Three versions of the apparatus have so far been developed (they are not being manufactured), for

ranges of test frequencies of 300-20 000 c/s, 15-1500 kc/s and 0.1-10 Mc/s. The frequencies quoted in the explanatory account that follows relate to the 15-1500 kc/s version, a photograph of which appears in *fig. 1*.

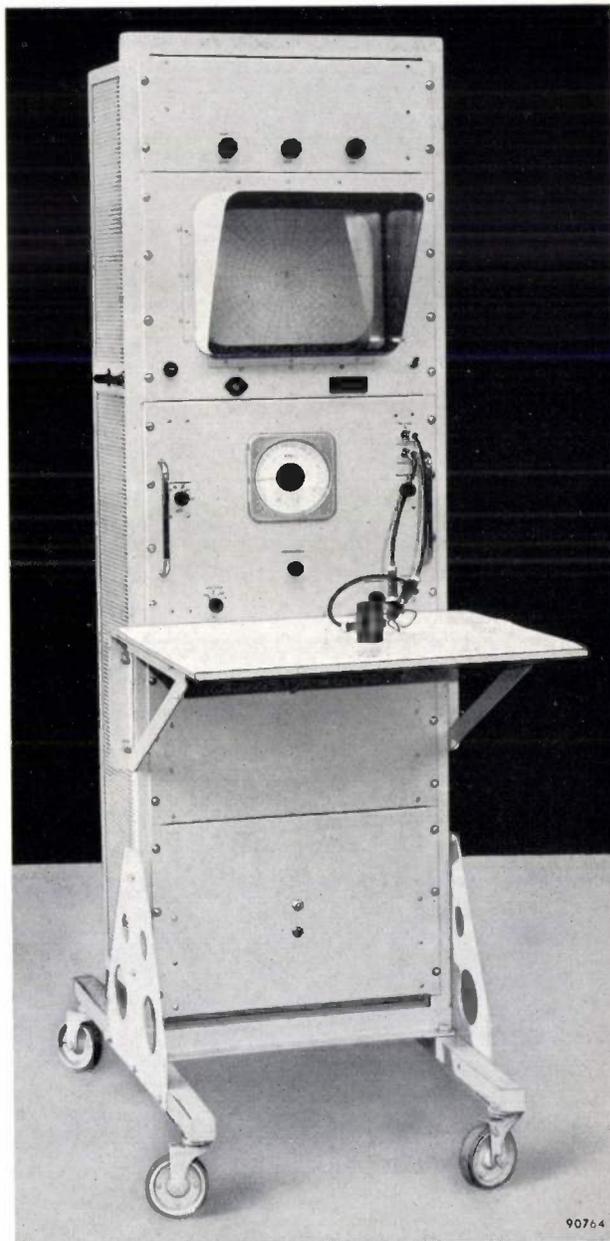


Fig. 1. Apparatus for the display of the complex ratio between currents and voltages at varying frequencies. The polar diagram is traced on the screen of a CRT having a long-persistence phosphor. The object under test seen on the table is a piezo-magnetic vibrator (pre-magnetized ferroxcube core with winding). The apparatus shown is suitable for test frequencies from 15 to 1500 kc/s.

<sup>1)</sup> See for example G. Thirup, An instrument for measuring complex voltage ratios in the frequency range 1-100 Mc/s, Philips tech. Rev. 14, 102-114, 1952/53.

<sup>2)</sup> C. M. van der Burgt, Ferroxcube materials for piezomagnetic vibrators, Philips tech. Rev. 18, 285-298, 1956/57.

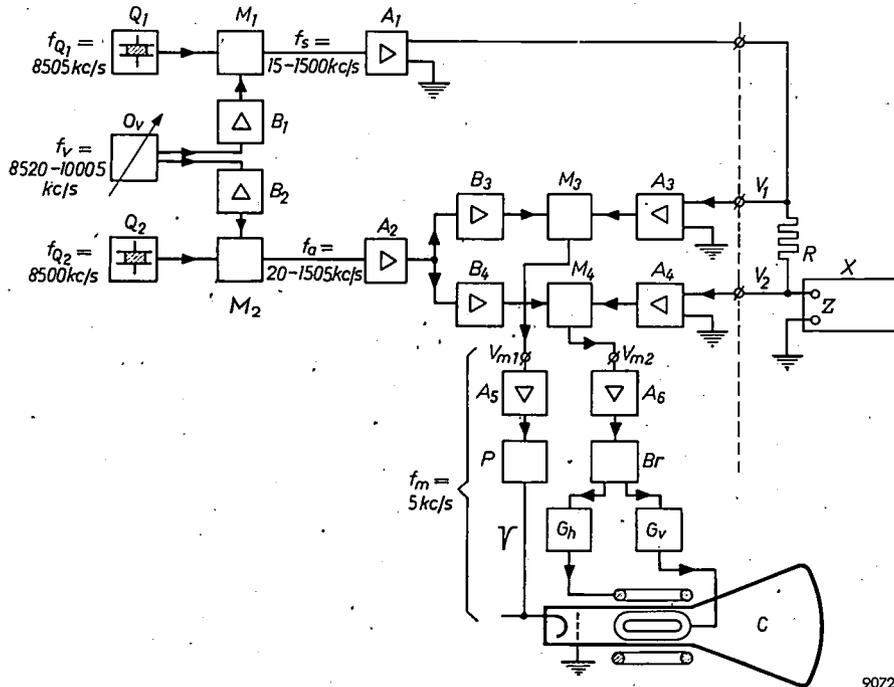
<sup>3)</sup> After this article had been written a paper by E. C. Pyatt (J. Brit. Instn. Rad. Engrs. 16, 563-567, 1956) came to our notice in which a similar type of instrument is described. Its relative frequency range is smaller, however, and it does not function automatically.

In the block diagram of *fig. 2*,  $Q_1$  and  $Q_2$  are crystal-controlled oscillators generating fixed frequencies of  $f_{Q1} = 8505$  kc/s and  $f_{Q2} = 8500$  kc/s. The voltages from  $Q_1$  and  $Q_2$  are fed to mixer stages  $M_1$  and  $M_2$  respectively.

Oscillator  $O_v$  supplies a frequency  $f_v$  which is variable between 8520 and 10 005 kc/s. The voltage from  $O_v$  is also fed to the mixer stages via buffer circuits  $B_1$  and  $B_2$ , these being isolating stages to prevent feedback. Now, we want to obtain from

the two signals it mixes, providing this stronger signal keeps within certain limits.

Mixer stage  $M_1$  delivers a signal of frequency  $f_v - f_{Q1}$  ( $= 15-1500$  kc/s) which we shall call  $f_s$ , the signal or test frequency. Amplifier  $A_1$  amplifies and passes on this signal to  $X$ , the apparatus under test, which might be an amplifier, a filter, a crystal vibrator or the like. Mixer stage  $M_2$  delivers a signal of frequency  $f_v - f_{Q2}$ , which is amplified by  $A_2$ . This auxiliary frequency  $f_a$  is always 5 kc/s



**Fig. 2.** Block diagram of the apparatus designed for a signal frequency range of 15-1500 kc/s.  $Q_1$  and  $Q_2$ : quartz-controlled oscillators.  $O_v$ : oscillator of adjustable frequency  $f_v$ .  $B_1, B_2, B_3$  and  $B_4$  denote buffer circuits,  $M_1, M_2, M_3$  and  $M_4$  mixer stages, and  $A_1, A_2, A_3, A_4, A_5$  and  $A_6$  amplifiers.  $P$ : pulse-shaper.  $Br$  is a bridge circuit, one output voltage of which has a  $45^\circ$  lead and the other a  $45^\circ$  lag with respect to the voltage on its input.  $G_h$  and  $G_v$  generators of the deflecting currents for cathode-ray tube  $C$ .  $X$  test object.  $R$  series resistor used when the impedance  $Z$  of test object  $X$  has to be measured ( $R \gg |Z|, Z \propto V_2/V_1$ ).  
 $f_v - f_{Q1} = f_s$ , the test or signal frequency.  $f_v - f_{Q2} = f_a$ , the auxiliary frequency.  
 $f_a - f_s = f_m$ , the intermediate frequency, equal to 5 kc/s.  
 The voltage delivered by  $A_1$  can be varied from zero to 1.5 V. The cable connecting  $A_1$  and  $X$  has a characteristic impedance of 135  $\Omega$  (this being the internal resistance of  $A_1$ ). In order that the pulse from  $P$  should be sufficiently strong,  $V_1$  must be approx. 5 mV. It is necessary that  $V_2$  should have the same value if the circle on the screen is to have a radius of 10 cm.

these mixer stages a signal whose amplitude is as nearly constant as possible, but the voltage from oscillator  $O_v$  inevitably varies rather considerably with the frequency  $f_v$ . This being so, it has been arranged that the buffer circuits deliver a voltage considerably higher than the constant voltages from  $Q_1$  and  $Q_2$ ; the reason for this is that the conversion transconductance of a mixer tube is more or less independent of the amplitude of the stronger of

higher than  $f_s$ , for the difference between  $f_{Q1}$  and  $f_{Q2}$  is 5 kc/s. We shall call this constant difference frequency  $f_m$ , the intermediate frequency.

From the test object  $X$  two voltages  $V_1$  and  $V_2$  are obtained that are in phase with and proportional in amplitude to the quantities whose complex ratio has to be measured. (In *fig. 2* it is assumed that an impedance  $Z$  is to be measured. A resistance  $R \gg |Z|$  is connected in series with  $Z$ , so that the

resulting current is practically independent of the impedance.  $V_2$  is the voltage across  $Z$ ;  $V_1$  is the voltage across  $R + Z$  and is therefore a measure of the current through them.) Voltages  $V_1$  and  $V_2$  are fed to the mixer stages  $M_3$  and  $M_4$  respectively via amplifiers  $A_3$  and  $A_4$ . The two voltages are mixed with a signal coming from amplifier  $A_2$  via the buffer circuits  $B_3$  and  $B_4$ . This auxiliary signal has the frequency  $f_a$  which, as stated above, exceeds the test frequency  $f_s$  of  $V_1$  and  $V_2$  by the constant amount of  $f_m = 5$  kc/s. The mixer stages produce two voltages  $V_{m1}$  and  $V_{m2}$  having the intermediate frequency. As can easily be demonstrated (see page 103 of the article cited in footnote<sup>1</sup>), these i.f. voltages have the same complex ratio as the h.f. voltages  $V_1$  and  $V_2$ , provided the mixer stages  $M_3$  and  $M_4$  have identical properties.

Continuing to follow the block diagram, we see that i.f. voltage  $V_{m2}$ , having been amplified in  $A_6$ , is applied to bridge circuit  $Br$  and there converted into two equal voltages, one of which has a  $45^\circ$  lag and the other a  $45^\circ$  lead with respect to  $V_{m2}$ . The two voltages control generators  $G_h$  and  $G_v$ , which supply the horizontal and vertical currents for magnetic deflection in the cathode ray tube  $C$ . The deflecting currents are sinusoidal and  $90^\circ$  out of phase. A circle would therefore be traced on the screen of the tube, were it not for the fact that the grid voltage is adjusted to a value such that in the normal state the electron beam is suppressed. The radius of the circle is proportional to the amplitude of  $V_{m2}$ , and hence to that of  $V_2$ .

The voltage  $V_{m1}$  is amplified in  $A_5$ ; by clipping its peaks and differentiating it, pulse-shaper  $P$  turns the amplified voltage  $V_{m1}$  into a pulse train. The pulse train controls the cathode ray tube in such a way that the beam suppression is lifted intermittently: a point on the circle then becomes visible. Since voltages  $V_{m1}$  and  $V_{m2}$  have the same frequency (namely  $f_m = 5$  kc/s), one point of the circle becomes visible on the screen. Which point becomes visible depends on the phase difference between  $V_{m1}$  and  $V_{m2}$ , and hence on the phase difference  $\varphi$  between  $V_1$  and  $V_2$ . The radius of the circle (i.e. the distance from the visible point to the centre) is a measure of the ratio between the amplitudes of  $V_1$  and  $V_2$ . Phase angle and radius can be read off on a grid of polar coordinates in front of the screen of the CRT.

As long as the frequency  $f_v$  of oscillator  $O_v$  is constant, the test frequency  $f_s$  has a fixed value and only one point of the vector diagram will be seen on the screen. If  $f_v$  is changed,  $f_s$  changes too, and the complex ratio being investigated generally

takes on a different value, and consequently a point of a different circle appears on the screen. If  $f_s$  is made to pass through a certain range of frequencies the spot will trace out a curve representing the value of the complex quantity throughout that particular frequency range.

As a rule, the first stage in the measurement of a complex quantity is to cause  $f_s$  to pass from its lowest to its highest value by slowly varying  $f_v$  by manual adjustment. It is then easy to see from the screen values of  $f_s$  (read on an instrument scale) at which any interesting phenomena occur. Resonance shows up as an ellipse or a circle such as may be seen in fig. 3. Even two resonances occurring close together — one of which is easily overlooked in measurements employing the usual bridge methods — are immediately recognizable.

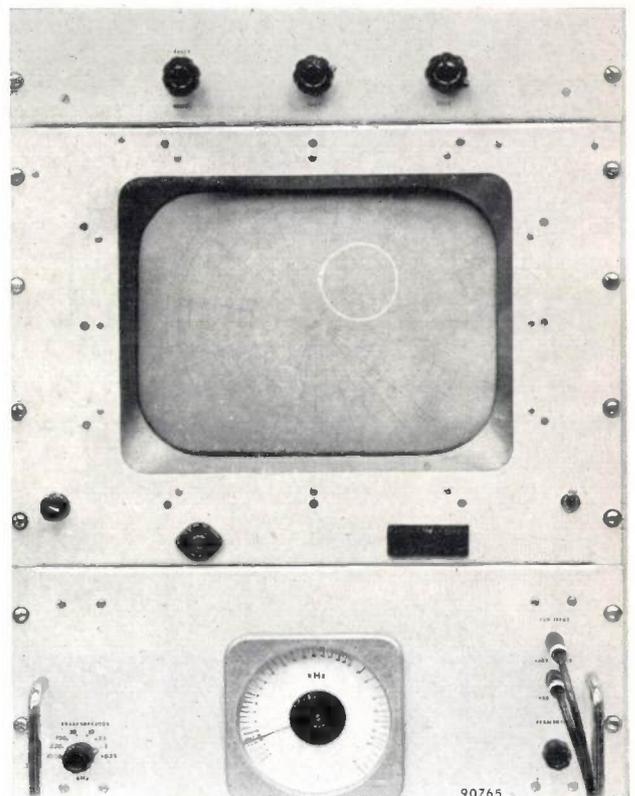


Fig. 3. On the screen can be seen one of the resonance circles of a piezomagnetic vibrator, occurring at a test frequency of about 35 kc/s. On the lower panel (left) is the control by which  $f_v$  (and hence  $f_s$  and  $f_a$ ) is varied;  $f_s$  can be read off on the frequency scale (centre).

If it is desired to give closer study to the behaviour of the test object in the vicinity of such a frequency, an arrangement is put into operation that causes frequency  $f_s$  to swing about the set value by an adjustable amount and in an adjustable tempo. For this purpose use is made of the fact that the reversible permeability (i.e. the permeability in

relation to weak alternating fields) of ferroxcube almost saturated by a static magnetic field is dependent on the strength of that field. Part of the inductance in oscillator  $O_v$  is wound on a ferroxcube core; this is pre-magnetized by a direct current, and on this direct current an alternating current of adjustable amplitude and adjustable low frequency can be superimposed. In this way periodic fluctuation is set up in the reversible permeability of the ferroxcube core, and consequently in the frequency  $f_v$ , and in frequencies  $f_s$  and  $f_a$  as a further consequence. The sweep of  $f_s$  increases in accordance with the amplitude of the superimposed alternating current. The sharper the resonance of the test object (i.e. the higher its  $Q$ ), the slower must be the rhythm in which  $f_s$  changes, in order that spurious responses near resonance be avoided. For the same reproducibility of the diagram and for the same frequency sweep (the width of the frequency range to be investigated is often determined by quantities other than the  $Q$ ) the rate at which  $f_s$  fluctuates must be made inversely proportional to the square of  $Q$ . The maximum rate of swing is in fact proportional to  $(f_s/Q)^2$ , as has been demonstrated by Siewert and Just<sup>4</sup>. The apparatus for a test frequency range of 15-1500 kc/s can be used for  $Q$  values up to about 1000 at 15 kc/s and for proportionally higher  $Q$  values at higher frequencies. In this version of the instrument the frequency sweep (i.e. the maximum deviation from the central frequency) is adjustable in steps from 0.15 to 500 kc/s and the frequency of the superimposed alternating current (and hence the recurrence frequency of the screen image) is adjustable in steps from  $1/30$  to 10 c/s. On account of this very low recurrence frequency, the CRT has a screen with a long-persistence phosphor.

The oscillograph method permits the measurement of complex quantities with an accuracy of within 5%. This means that the true end of the vector representing the complex quantity lies within a radius of 0.5 cm from the observed point when the vector has the maximum length of 10 cm. Hence the modulus can be measured with an accuracy of within 5% and the phase angle with an accuracy of within  $1/20$  radian, or approx.  $3^\circ$ , for a full-scale reading. For smaller readings both modulus

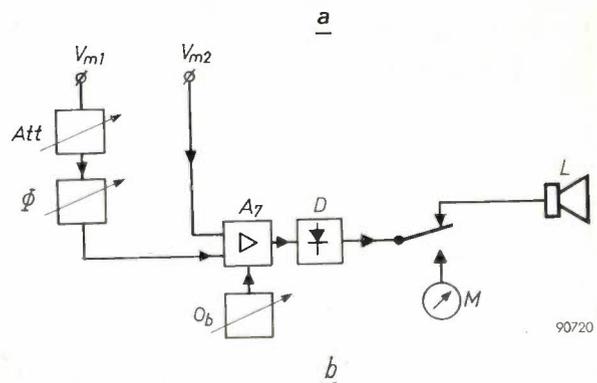
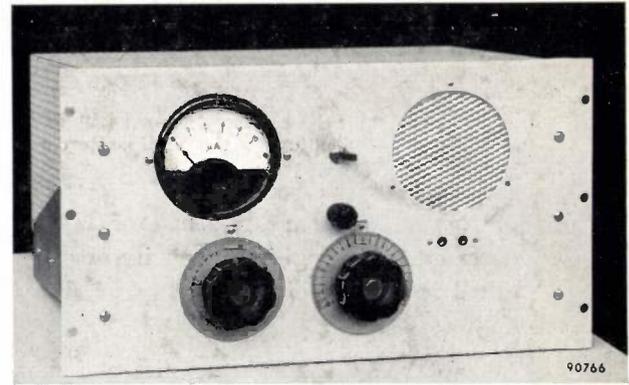


Fig. 4. a) Auxiliary unit with the aid of which the oscillograph method can be replaced by a null method<sup>1)</sup>, should greater accuracy be required. b) Block diagram of the auxiliary unit shown in (a). *Att*: calibrated attenuator.  $\Phi$ : calibrated phase-shifter.  $A_7$ : amplifier.  $O_b$ : beat oscillator.  $D$ : detector. Either meter  $M$  or loudspeaker  $L$  can be used for null indication. For the I.F. voltages  $V_{m1}$  and  $V_{m2}$ , refer to fig. 2.

and phase angle are given with proportionally less accuracy.

If higher accuracy is desired, use can be made of the method described in the article cited in footnote<sup>1)</sup>, and the auxiliary unit shown in fig. 4a added to the apparatus. In this way the oscillograph method is replaced by a null method. The block diagram of the auxiliary unit appears in fig. 4b. From the I.F. voltage  $V_{m1}$  a voltage is derived that is made equal in amplitude and phase to  $V_{m2}$  by means of a calibrated attenuator and a calibrated phase shifter. Equality between the two voltages is detected with the aid of an amplifier, a beat oscillator, a detector and some form of null-indicator. This last may be a meter or a loudspeaker. This method makes it possible to measure a modulus with an accuracy of within 0.2 dB (2%) and a phase difference with an accuracy of within  $2^\circ$ , irrespective of amplitude, provided this is not too low.

M. J. van SCHAGEN.

<sup>4)</sup> W. Siewert and G. Just, Ein Kurvenschreiber für Quarzimpedanzen, Telefunken Zeitung 28, 45-51, 1955.

## ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk \* can be obtained free of charge upon application to the Philips Research Laboratory, Eindhoven, Netherlands.

**R 306:** H. G. van Bueren: Influence of lattice defects on the electrical properties of cold-worked metals (Philips Res. Rep. 12, 1-45, 1957).

The first chapter of this thesis contains a review of the physical properties of dislocations, vacancies and interstitial atoms in the face-centred cubic metals copper, silver and gold. The mechanism of plastic deformation of such metals is then considered theoretically in order to obtain numerical data on the concentrations of the various imperfections present in the metal after cold-work. Measurements of electrical resistivity of slightly deformed copper and silver wires are presented and compared with the theory, thereby throwing light on several details of the deformation mechanism. By measuring the magnetoresistance of the strained wires, it is possible to separate the influences of dislocations from that of point imperfections. This, combined with observations of the thermal recovery of electrical conductivity after cold-work, a recovery characterized by four separate recovery stages, yields information on the diffusive as well as on the electron-scattering properties of the three types of lattice imperfections.

**R 307:** J. Haantjes and G. J. Lubben: Errors of magnetic deflection, I. (Philips Res. Rep. 12, 46-68, 1957).

This paper deals with the errors caused by magnetic deflection of the electron beam in a cathode-ray tube. It appears that image distortion, astigmatism, curvature of the image field and coma may occur, corresponding to the errors of the third order in conventional optics. It will be shown that in most cases one may influence and even eliminate these errors by a suitable distribution of the deflecting field. The only error which can never be eliminated by a special field distribution is the error known as curvature of the image field. This error would not appear if the screen face of the tube coincided with the image field. However, the curvature of the image field is such that it is generally impossible to

fulfil this condition. Therefore this error has to be taken into account in the design of cathode-ray tubes for television purposes.

**R 308:** K. F. Niessen: Spontaneous magnetization and magnetic susceptibilities of an antiferromagnetic with foreign ions in both sublattices (Philips Res. Rep. 12, 69-81, 1957).

The spontaneous magnetization is calculated as a function of the temperature for an antiferromagnetic, in the two sublattices of which a different but relatively small number of original ions are assumed to be replaced by ions of another metal, carrying another magnetic moment. For a real calculation the Néel temperature of the material must be measured and also the Néel temperature of the original antiferromagnetic (in which the above replacement has not yet taken place). From the ingredients used for the preparation of the material the sum of the foreign ions in the two sublattices will be known, but not their ratio. This ratio, however, can be derived as a function of the temperature from the spontaneous magnetization. Moreover, we determine the change in the parallel and perpendicular susceptibilities, due to the above replacement of original by foreign ions.

**R 309:** B. H. Schultz: Storage of injected carriers at surfaces of germanium (Philips Res. Rep. 12, 82-96, 1957).

The relative number of particles at the surface in the presence of an injection is calculated with the aid of standard space-charge theory. The effect of this "surface storage" on the relaxation of photoconductance is discussed. Measurements have been made of  $\tau_{rel}$  of thin *n* and *p* type germanium rods as a function of their thickness in different atmospheres. The results are in qualitative agreement with the theory. For rods with large surface storage,  $\tau_{rel}$  increases if the thickness is reduced and becomes very sensitive to the injection level.