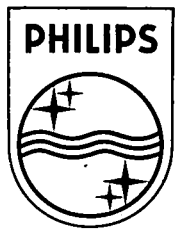
PHILIPS TECHNICAL REVIEW

12731-36.



13 DEC. 1958

Bibl. Nat. Lab.

N. V. PHILIPS' GLOEILAMPENFABRIEKEN

VOLUME 19

1957/1958

PHILIPS TECHNICAL REVIEW

A publication of the Research Laboratories of N.V. Philips' Gloeilampenfabrieken, Eindhoven, Netherlands. Twelve issues per year and an average of 32 pages per issue. The Review is published in four editions: English, Dutch, French and German, all identical in contents.

Editors

S. Gradstein, Editor-in-chief
D. M. Duinker, Deputy editor
M. Beun
C. M. Hargreaves
J. W. Miltenburg

All rights reserved by N.V. Philips' Gloeilampenfabrieken, Eindhoven, Netherlands. Articles may be reproduced in whole or in part, provided that the editors are first notified and that the source "Philips Technical Review" is mentioned in full. Photographs and drawings for this purpose are available on request.

CONTENTS
VOLUME 19, 1957/1958

LIGHT

	Page
Electroluminescence and image intensification, by G. Diemer, H. A. Klasens and P. Zalm	1- 11
Applications of the interference of light in thin films, by P. M. van Alphen	59- 67
The function of additives in tungsten for filaments, by J. L. Meijering and G. D. Rieck	109-117
A luminous frame around the television screen, by J. J. Balder	156-158
An open-air laboratory for road lighting, by I. Hamming and J. F. T. van Heemskerck Veeckens	202-205
A mirror-condenser lamp for 8-mm projectors, by P. M. van Alphen and M. Bierman	233-235
The double-flux "TL" lamp, a fluorescent lamp of high output per unit length, by H. J. J. van Boort and D. Kolkman	333-337

TELECOMMUNICATIONS

	Page
The junction transistor as a network element at low frequencies, by J. P. Beijersbergen, M. Beun and J. te Winkel, II. Characteristics and h parameters	15- 27
II. Equivalent circuits and dependence of the h parameters on operating point	98-105
Pre-magnetization of the core of a pulse transformer by means of ferroxdure, by H. G. Bruijning and A. Rademakers	28- 37
Audio amplifiers with single-ended push-pull output, by J. Rodrigues de Miranda	41- 49
Applications of the interference of light in thin films, by P. M. van Alphen	59- 67
"Frena", a system of speech transmission at high noise levels, by F. de Jager and J. A. Greefkes	73- 83
Fundamentals of colour television, by F. W. de Vrijer	86- 97
A transmitting triode for frequencies up to 900 Mc/s, by P. J. Papenhuijzen	118-128
A 4000 Mc/s wide-band amplifier using a disc-seal triode, by J. P. M. Gieles	145-156
A luminous frame around the television screen, by J. J. Balder	156-158
Dispenser cathodes,	
I. Introduction, by A. Venema	177-179
II. The pressed cathode, by R. C. Hughes and P. P. Coppola	179-185
III. The impregnated cathode, by R. Levi	186-190
A radar sonde system for upper air measurements, by N. E. Goddard and H. A. Dell	258-263
A method of sealing the window and cone of television picture-tubes, by A. H. Edens	318-323
The projection of colour-television pictures, by T. Poorter and F. W. de Vrijer	338-355

SOUND

A transistor hearing aid, by P. Blom and P. Boxman	130-139
Electronic music, by H. Badings and J. W. de Bruyn	191-201
Fabrication of sapphire and diamond styli for gramophone pick-ups, by A. W. Ploegsma	324-326

X-RAYS AND NUCLEAR PHYSICS

Chladni's figures on the vibrating capacitor of a synchrocyclotron, by B. Bollée	84- 85
A simple apparatus for contact microradiography between 1.5 and 5 kV, by B. Combée and A. Recourt	221-233
Monitoring, control and safety equipment for a nuclear reactor of the swimming-pool type,	
I. General description, by M. van Tol	245-257
II. Further description of certain component units, by F. E. L. ten Haaf, G. Klein and F. J. Schijff	273-285
Dosimetry of the very weak X-radiation generated in television receivers and X-ray diffraction apparatus, by W. J. Oosterkamp, J. Proper and J. J. F. de Wijk	264-267
An X-ray diffraction tube with rotating anode for 10 kW continuous loading, by W. J. H. Beekman, A. Verhoeff and H. W. van der Voorn	314-317

MEASURING EQUIPMENT

Stroboscopic operation of photomultiplier tubes, by C. F. Hendee and W. B. Brown 50- 58

Low-voltage oscilloscope tubes, by F. de Boer and W. F. Nienhuis 159-164

Monitoring, control and safety equipment for a nuclear reactor of the swimming-pool type,
I. General description, by M. van Tol 245-257
II. Further description of certain component units, by F. E. L. ten Haaf, G. Klein and F. J. Schijff 273-285

A radar sonde system for upper air measurements, by N. E. Goddard and H. A. Dell 258-263

Dosimetry of the very weak X-radiation generated in television receivers and X-ray diffraction apparatus, by W. J. Oosterkamp, J. Proper and J. J. F. de Wijk 264-267

MATERIALS AND THEIR RESEARCH

The sub-microscopic structure of "Ticonal" G magnet steel, by H. B. Haanstra, J. J. de Jong and J. M. G. Smeets . . 11- 14

Pre-magnetization of the core of a pulse transformer by means of ferroxdure, by H. G. Bruijning and A. Rademakers 28- 37

The function of additives in tungsten for filaments, by J. L. Meijering and G. D. Rieck 109-117

Dispenser cathodes,
I. Introduction, by A. Venema . . . 177-179
II. The pressed cathode, by R. C. Hughes and P. P. Coppola . . . 179-185
III. The impregnated cathode, by R. Levi 186-190

Crystal-oriented ferroxplana, by A. L. Stuijts and H. P. J. Wijn 209-217

An improved method of cleating, by H. E. Deelman 268-269

Direct observation of Weiss domains by means of the Faraday effect, by C. Kooy 286-289

Barium getter films, by J. J. B. Fransen and H. J. R. Perdijk 290-300

Paramagnetic resonance, by J. S. van Wieringen 301-313

The generation of dislocations by thermal stresses, by P. Penning 357-364

MISCELLANEOUS

Electroluminescence and image intensification, by G. Diemer, H. A. Klasens and P. Zalm 1-11

A high-precision lathe headstock, by L. M. Leblans 68- 69

Chladni's figures on the vibrating capacitor of a synchrocyclotron, by B. Bollée . 84- 85

An application of television for the discovery of variable stars, by J. Borgman 140-142

The formulation of pesticides, by W. Duyfjes 165-176

Analysis of the gaseous contents of sealed cathode-ray tubes with the aid of the omegatron, by J. Peper 218-220

The Foucault pendulum in the United Nations Building in New York, by J. A. Haringx and H. van Suchtelen . . . 236-241

An improved method of cleating, by H. E. Deelman 268-269

Barium getter films, by J. J. B. Fransen and H. J. R. Perdijk 290-300

A method of sealing the window and cone of television picture-tubes, by A. H. Edens 318-323

Fabrication of sapphire and diamond styli for gramophone pick-ups, by A. W. Ploegsma 324-326

VOLUME 19, 1957/58, No. 1 pp. 1-40 Published 27th July 1957

Philips Technical Review

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

ELECTROLUMINESCENCE AND IMAGE INTENSIFICATION

by G. DIEMER, H. A. KLASENS and P. ZALM.

535.376:621.383.2:
621.3.032.36

The phenomenon of electroluminescence, which was discovered in 1936, has so far not fulfilled the expectations it aroused as regards its usefulness as a light source: although it can now be exploited in "luminous panels" producing light with reasonable efficiency (up to about 10 lumens/watt), it seems that such panels, supplied from the mains, would have to have very large surface areas to give the luminous flux normally required for lighting a room. On the other hand, the fact that local differences of brightness can be produced in a luminous panel in a simple fashion has opened up unexpected perspectives. On this fact are based the solid-state image intensifiers that are arousing increasing interest.

Introduction

An ideal light source would be one which converted all the energy supplied to it (e.g. heat or electrical energy) into light of the desired spectral composition (e.g. that of daylight) in the simplest possible way. Up to now the ideal is most nearly approached by the fluorescent lamp, a combination of the gas discharge and the phenomenon of fluorescence. There nevertheless remains a demand for a technically simpler process of energy conversion, one for example which would do away with the need for a vacuum-tight envelope (as is required by a gas discharge) and which would not involve corrosion or any other kind of ageing of the light source. The discovery of the phenomenon of electroluminescence[1]) by the Frenchman Destriau in the 'thirties does seem to have been a step forward on the road to direct conversion of electrical energy into light without ageing of the source, although we must admit that, despite recent advances in this field, light sources based on this phenomenon still do not seem to be of practical use for general lighting.

A light source based on electroluminescence is simple enough: it merely consists of say, a layer of suitably prepared substance (usually zinc sulphide), coated on a transparent base, and placed in an electric field. At voltages considerably lower than the breakdown voltage of the layer, the passage of current through the zinc sulphide results in the emission of light, no physical or chemical changes arising in the substance, and the heat developed is barely perceptible. An electroluminescent layer of this kind is conveniently named a *luminous panel.*

The photometric quantity used for measuring the light output of a source such as this, which radiates from an extended area, is the luminous emittance, this being the total number of lumens radiated from one side of the plane per square metre of its area. Luminous panels made so far have only had a moderate emittance (1 m^2 of an electroluminescent panel may at most emit as many lumens as a 200 W incandescent lamp) and not particularly high efficiency; for the time being, therefore, the employment of the luminous panel as a light source is likely to be confined to such special cases as the lighting of instrument dials, radio receiver scales, clock faces, etc. However, the fact that local differences of emittance can be produced in a simple way is of great interest, and attempts are being made to exploit this in solid-stage image intensifiers,[2]) to which we shall devote a word at the end of this article.

[1]) G. Destriau, J. Chim. phys. **33**, 587-625, 1936; Phil. Mag. (7) **38**, 700-739, 774-793, 880-888, 1947. For a survey of the subject of electroluminescence, especially that of zinc sulphide powders, see P. Zalm, Philips Res. Rep. **11**, 353-399 and 417-451, 1956 (Nos. 5 and 6); G. Destriau and H. F. Ivey, Proc. Inst. Rad. Engrs., Solid-state materials issue, December 1955.

[2]) See for example G. Diemer, H. A. Klasens and J. G. van Santen, Philips Res. Rep. **10**, 401-424, 1955.

The article will be mainly concerned with a description of the complicated mechanism of light production in a luminous panel. We shall first discuss one or two important properties of the device.

Construction and general behaviour of a luminous panel

A diagram showing the construction of a luminous panel appears in *fig. 1*. The panel consists of a glass base to which a transparent electrode is applied,

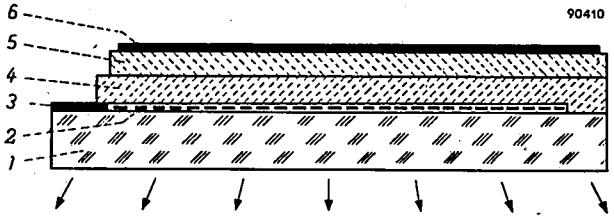


Fig. 1. Diagram showing construction of an electroluminescent cell or "luminous panel". *1* glass plate (base). *2* very thin, transparent, conducting coating of tin oxide. *3* metal contact strip. *4* layer of zinc sulphide suspended in a transparent lacquer (approx. 25 microns thick). *5* reflector layer of white lacquer (approx. 20 microns thick). *6* metallic layer, constituting the second electrode. Between this last and the tin oxide coating is applied an alternating voltage of several hundred volts at a frequency in the audio range (50-10 000 c/s). A second glass plate is usually glued on to electrode *6* to protect the panel from damage due to ingress of moisture.

e.g. a very thin layer of tin oxide sprayed on at high temperature. On top of the transparent electrode is applied a layer of electroluminescent substance, consisting of a powder (zinc sulphide) prepared in a special way and suspended in a transparent insulating lacquer. This second layer is about 25 microns thick. In order that the light produced in it should as far as possible be reflected

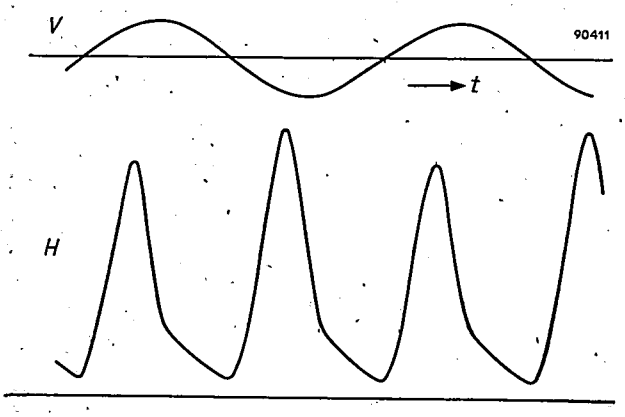


Fig. 2. Sinusoidal alternating voltage V and the electroluminescent effect (emittance H) it produces, as functions of time. A flash is produced every half-period. The frequency of the voltage depicted here was 65 c/s.

outwards via the transparent electrode, it is generally given a coating of white lacquer before the panel is backed with a second (metallic) electrode. Thus the panel as a whole represents a parallel plate condenser.

If we apply a sufficiently high alternating voltage to the panel, a flash of light is seen to be emitted during each half-period of oscillation (*fig. 2*). Sometimes, depending on the nature of the electroluminescent substance and the frequency of the voltage applied, a second and less intense flash is observed. Provided the frequency is neither too

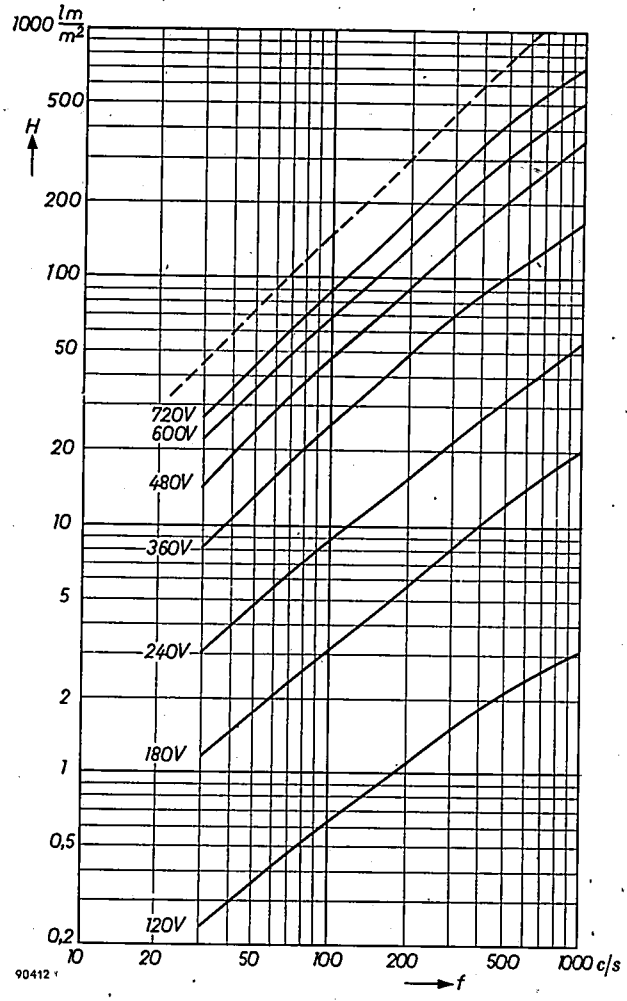


Fig. 3. If the r.m.s. value of voltage V applied to the cell be kept constant, emittance H increases near enough directly proportional to frequency f. (The curves are roughly parallel to the broken line at 45°.)

low (not below 10 c/s) nor too high (not above 1000 c/s), the amount of light emitted in each flash is found to be more or less independent of it; hence H, the emittance of the panel averaged out over a period of time, becomes greater for the same voltage at a higher frequency (being roughly in direct proportion to it). This can be seen from *fig. 3*.

With frequency kept constant, changes in the voltage V are found to produce very large changes in emittance. At low values of voltage the emittance increases sharply with increasing voltage, less sharply at higher values. A simple formula relating the emittance H and the voltage V is found by plotting the logarithm of the emittance (which is quite a good measure for our subjective impression of brightness) against the reciprocal of the root of the voltage (see *fig. 4*). This gives a straight line with a negative slope:

$$\ln H = \ln H_0 - \frac{\text{const.}}{\sqrt{V}} \quad . \quad . \quad . \quad . \quad (1)$$

In order therefore to obtain a high emittance the panel must be operated at a high frequency and at a voltage as high as its breakdown strength will permit.

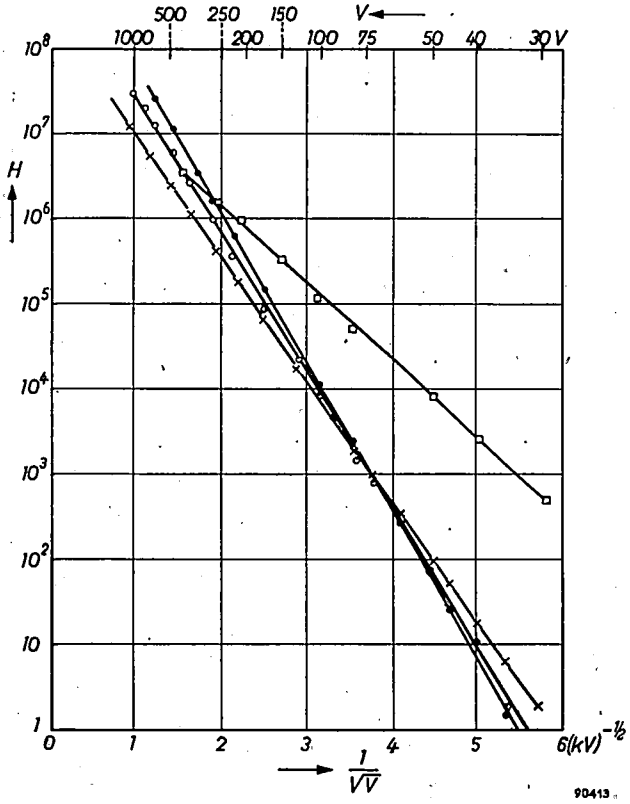


Fig. 4. The logarithm of the emittance H is a linear function of $1/\sqrt{V}$. The above curves were obtained experimentally, with various zinc sulphide powders, the H scale being in arbitrary units.

Mechanism of electroluminescence

The functioning of a luminous panel differs from other familiar kinds of luminescence, such as cathodoluminescence (as in television picture tubes) and photoluminescence (as in fluorescent lamps), only in the way in which the excitation of the luminescent substance takes place [3]). In cathodoluminescence, excitation is brought about by bombardment with a beam of fast electrons which, in penetrating the substance, raise certain bound electrons to a higher energy state; the bound electrons are thereby rendered capable of emitting light quanta $h\nu$ (h = Planck's constant = 6.62×10^{-34} joule sec; ν = frequency of the light = number of vibrations per second) on reverting to the ground state. In photoluminescence the same kind of excitation is caused to take place in the substance by its absorption of short-wave radiation, such as ultra-violet quanta, falling upon it. It seems that electroluminescence, as in a luminous panel, can best be understood by supposing that mobile electrons are present in the electroluminescent substance, that these electrons are able to acquire energy from the electric field and that, when they have sufficient energy, are able to bring bound electrons in the crystal into an excited state by colliding with them (impact excitation).

We shall now discuss the mechanism of excitation by impact in somewhat more detail.

Let us take ZnS as an example of an electroluminescent substance. A crystal of this substance, taken as a whole, is electrically neutral; electrical neutrality will also be found to obtain within each small element of its volume. Let us suppose the crystal to be coated at opposite ends with electrodes and that a strong electric field F is set up in it by a voltage on the electrodes. If now by some means a few electrons are introduced into the ZnS crystal, they will move freely in the direction of the field, as they would in a gas discharge, travelling an average distance of, say, l. In this way they are accelerated by the field up to an energy of eFl before giving up their acquired energy in collisions (e = charge of the electron = 1.6×10^{-19} coulomb). In an ideal crystal with its atoms or ions arranged with complete regularity, the mean free path l may amount to many interatomic distances, despite the density of the solid. (This means that the chance of collision is very much less than in a gas of the same density, where there is no regular arrangement.) In ZnS, for example, $l \approx 100$ Å, this being roughly equivalent to 40 interatomic distances. Collision being a matter of chance, there will be some electrons that travel a much longer distance than 100 Å before they collide. The theory of probability gives the following relation for n, the

[3]) For a general introductory discussion of fluorescent substances (phosphors) and the mechanism of fluorescence, see for example F. A. Kröger, Philips tech. Rev. 6, 349-358, 1941.

number of electrons having a free path of x instead of the mean free path l:

$$\ln n = \ln n_0 - \frac{x}{l}, \quad \ldots\ldots \quad (2)$$

where n_0 is a constant proportional to the total number of electrons. Let the energy necessary for the excitation process described above be E_v. To acquire this amount of energy the electron must travel a distance x, without meeting an obstacle, such that

$$eFx = E_v,$$

i.e.

$$x = \frac{E_v}{eF}. \quad \ldots\ldots \quad (3)$$

Inserting this in (2), we obtain the following for the number of successful collisions:

$$\ln n_x = \ln n_0 - \frac{E_v}{eFl}. \quad \ldots\ldots \quad (4)$$

If we assume that there is a certain fixed chance of the bound electron giving rise to the emission of light on reverting to the ground state, then in the stationary state the rate of excitation must be proportional to the rate of emission. We can therefore compare the excitation formula, eq. (4), with eq. (1), the formula for the emittance found experimentally. On doing so, we see that both portray the same kind of dependence on voltage provided that the field strength F that accelerates the electrons is proportional to the *root* of the applied voltage. This seems strange at first sight because, after all, the field strength in a plate condenser (in which the field is homogeneous) is nothing other than the applied voltage divided by the distance between the electrodes, and is therefore proportional to the voltage itself, and not to its root.

We must seek a connection between this apparent inconsistency and another remarkable fact: perceptible emission of light begins at voltages far below the breakdown voltage of pure zinc sulphide (at voltages 10 to 100 times lower, in fact), whereas it is only when the breakdown voltage has been reached that the field strength — as calculated for a homogeneous field — has increased to the point where the high-energy collisions necessary for excitation become really numerous [4]). This paradox can be resolved by assuming that the field in the crystal is *not* homogeneous and that a considerable part of the voltage applied to it acts across a very thin layer, wherein it produces a local field strength much higher than the average. In this very thin layer the electrons would have a much greater chance of collecting sufficient energy for excitation in the time elapsing before their next collision.

A "barrier" of this kind could indeed arise, on the cathode side of the crystal, say, in the following way. We are assuming that in the process of excitation electrons are liberated in the crystal, being freed by collision, so that they are able to move in the direction of the anode. If an electron is removed from an electrically neutral volume element, an equal positive charge will be left over. By adding certain foreign ions (in a concentration of 10^{-3} or 10^{-4}) having a lower valency than the zinc ions belonging to the lattice — copper ions, for example — the positive charge thus formed will remain bound in the vicinity of a nearby copper ion. At the same time, the excitation energy is stored in these bound positive charges: emission of light takes place when, on reversal of the field, the displaced electron recombines with the positive charge. If at one place in the crystal a number of positive charges be created (and the chance of this happening is greatest near the cathode, as we shall see later), then the field strength ceases to be constant throughout the crystal; the region of the positive space charge has a much greater density of lines of force than the average, because these lines of force all run from the positive space charge to the negative charge induced on the electrode nearby (see *fig. 5*). In this high-field region a high rate of excitation ensues in accordance with eq. (4), further increasing the space charge there and causing the field to contract ever more closely into a thin layer.

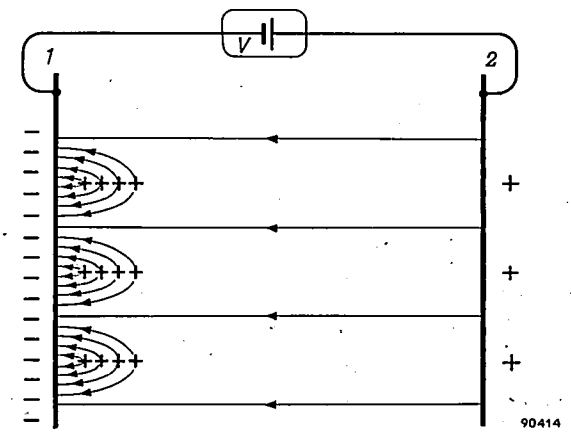


Fig. 5. In a zinc sulphide crystal provided with electrodes *1* and *2* connected to a voltage source V, a layer of positive space charge of certain thickness is set up on the cathode side of the crystal. The field strength in this layer, as indicated by the lines of force in the diagram, is much greater than the average field strength in the crystal.

[4]) This follows from eq. (4) if E_v be given the value 2.5 eV, i.e. the energy corresponding to a visible light quantum.

This explains the remarkable fact that emission of light occurs even at low voltages; it also allows us to explain why the field strength F, which according to eq. (4) is responsible for excitation, is proportional not to V, the applied voltage, but to its root. We assume that the accumulation of positive charges near immobile copper ions continues to the point where almost the whole of the voltage is acting across the space-charge layer and that, in conditions of maximum excitation, all the copper ions possess extra positive charges. If a higher voltage is applied to the crystal, there will be an increase in the total space charge Q under conditions of maximum excitation; but since all the copper ions already have excess charges, the increase in Q must be accompanied by a thickening of the space-charge layer, the depth of which in fact is proportional to Q. If we regard the space-charge layer as a charged capacitor with a plate separation of d (d being the thickness of the layer adjusted by a factor), we find that its capacitance C_b, which is proportional to $1/d$, decreases with increasing voltage. From the formula giving the charge on a capacitor:

$$Q = C_b \times V,$$

it follows that

$$V \propto d^2. \quad \ldots\ldots\ldots \quad (5)$$

However, F, the field strength in the barrier, is given by V/d, hence

$$F \propto d. \quad \ldots\ldots\ldots \quad (6)$$

Comparison of eqs. (5) and (6) shows that the field strength producing excitation is proportional to the root of the applied voltage, this bringing eq. (4) into line with eq. (1), the emittance law based on experimental observations.

Grains of zinc-sulphide powder suspended in a lacquer make no electrical contact with the electrodes and with one another. They therefore have to be provided with a very thin conductive skin or, to be more exact, with little conductive patches (to prevent short-circuiting from one side of the grain to the other). These "local electrodes" serve to collect the induced negative charge which, together with the positive space charge in the crystal, creates the strong field that causes excitation (see fig. 5). Moreover, even in the absence of an external voltage, there is a discontinuity in potential near the surface of the grain ("contact potential") due to the mere contact of the ZnS crystal and its conductive skin. This makes it understandable that excitation should tend to start in the surface of the grain on its cathode side, and that the subsequent formation of a positive space charge should cause the process to be gradually intensified until the stationary state is reached.

The conductive patches on the ZnS grains may be formed by adding to the zinc sulphide more copper than is soluble in it at a certain temperature; the remainder then separates out as a conductive compound of copper (probably copper sulphide). The presence of conductive patches on the surface of its grains is one of the most important characteristics distinguishing electroluminescent zinc sulphide from the normal luminescent kind.

As already stated, when an alternating voltage is applied to the panel the energy of excitation is released as light each time the field is reversed, the electrons being driven back to the positive charges and reverting to their ground state of energy ("recombination"). If an alternating voltage is suddenly applied at full strength to a virgin crystal, no light is emitted in the first half-period of oscillation. During the subsequent periods the recombination flash gradually becomes brighter until the stationary value is attained. This indicates that it is not possible during the first half-period to build up the space-charge barrier to the depth it ultimately reaches. If after some time the alternating voltage is withdrawn, and re-applied after a pause, the barrier builds up much more quickly than it does in a virgin crystal, even if the waiting interval has lasted several seconds; furthermore, a flash is now emitted during the first half-period. This proves that the recombination process taking place as the displaced electrons are swept back is not a complete one: the greater part of the positive space charge is left over each time, and that part recombines only very slowly after the field has been removed.

That light originates only in very small regions of the crystal is shown by microscopic examination. For a more extensive description and explanation of these and other details of the very complicated phenomenon of electroluminescence (the above-mentioned second flash in each period, for example), the reader is referred to the literature cited in footnote [1]).

So far we have been concerned solely with A.C. luminescence. Zinc sulphide phosphors also exist that emit light when a direct voltage is applied to them (provided arrangements are made for direct current to pass through the zinc sulphide). The emission of light is made possible by an activator which, when excited by impact, does not lose its electron but keeps it bound in a higher energy state.

In this way the electron is able to revert to the ground state, occasioning the emission of light, without the field being reversed. Manganese is an activator of this kind; in addition to manganese, however, it is necessary to introduce activators such as Cu into the ZnS in order to ensure that a space-charge barrier shall be formed.

The question might be posed as to where the first electrons come from that enable the excitation process to make a start, thereby allowing the barrier to build up. The answer cannot be given with certainty, but it is probable that these electrons come from the "local electrode" patches on the zinc sulphide. It is also possible that bound electrons occur in irregularities in the outermost atomic layers of the crystal itself and that these electrons can be released by the field acting in that region.

To conclude this explanation of the mechanism of electroluminescence it may be mentioned that a further light phenomenon exists which is frequently referred to as electroluminescence, and in which electrical energy in a solid is directly converted into light without first being converted into heat, but which is based on an entirely different mechanism. This is the phenomenon in which light is produced by the recombination of holes and electrons at a *p-n* junction [5]) when a voltage is applied to it in the forward direction. This kind of electroluminescence is found, for example, in SiC [6]), Ge [7]) and CdTe [8]). In a sense, it can be regarded as a reversal of the familiar and important photo-electric effect exhibited by *p-n* junctions. We shall not go further into the matter here.

Fabrication of the luminous panel

The foregoing will probably have made it clear that in every case an electroluminescent substance must also have fluorescent properties, in other words that it must be a phosphor [3]). Amongst all the many known phosphors, ZnS is eminently suitable as the basic material for electroluminescent purposes and there are two reasons for this. Firstly ZnS can be activated in such a way that electrons have a reasonably high mobility within it and are therefore capable of being accelerated up to sufficiently high energies. Secondly, ZnS so activated allows the space-charge barriers described above to build up easily. A further important point is that, in zinc sulphide, the Cu centres to which the positive charge is bound are stable in strong electric fields.

The conditions necessary for the chemical preparation of electroluminescent ZnS have been investigated in detail by Zalm [1]), making use of the earlier work of Kröger *et al.* [9]) on ZnS phosphors activated with copper. The solubility of Cu in ZnS is to a great extent influenced by the "coactivator" (Cl, Al, etc.); it is of importance that the concentration of Cu (expressed in gram-atoms) should be somewhat greater than that of the coactivator, to ensure that there is an excess of copper that can precipitate on cooling. Further, the concentrations of both are made sufficiently high to guarantee effective activation. As in other types of luminescence of ZnS activated with Cu, it is found that the product can be made to luminesce with a blue or green light according to the concentrations of activator and coactivator used. By further adding manganese an orange-yellow light can be obtained. By replacing sulphur by selenium in increasing proportions (that is, by taking mixed crystals of ZnS-ZnSe as basis for the powder), the light emitted by green or blue electroluminescent ZnS powders can be shifted into a continuous range of longer wavelengths. The addition of organic fluorescent substances capable of converting visible light of short wavelength into orange or red light makes it possible to produce panels giving a light of practically any desired colour.

A suspension of the electroluminescent powder is made in a synthetic lacquer which forms the binder; then, by means of known techniques such as spraying or "silk-screen printing", a homogeneous coating of the suspension about 35 μ thick is applied to a glass base previously rendered conducting by means of a tin-oxide coating. The lacquer should preferably have the following properties: 1) a high dielectric constant ε, to ensure that as much as possible of the alternating voltage acts across the ZnS grains (which themselves have a high ε); 2) a high breakdown voltage, so that the panel can be made to give high emittance values; and 3) a low moisture content (and also as moisture-*proof* as possible). Electrolysis may take place in an electroluminescent layer containing moisture, and this will shorten the life of the panel; moreover, moisture causes high dielectric losses that are non-productive of light. Urea formaldehyde resin was found to give a reasonably good compromise to these somewhat conflicting requirements.

---

[5]) See for example F. H. Stieltjes and L. J. Tummers, Philips tech. Rev. **17**, 233-246, 1955/56.

[6]) O. W. Lossew, Phys. Z. **34**, 397-403, 1933; C.R. Acad. Sci. U.R.S.S. **39**, 363-369, 1940.
K. Lehovec, C. A. Accardo and E. Jamgochian, Phys. Rev. **89**, 20-25, 1953.

[7]) J. R. Haynes and H. B. Briggs, Phys. Rev. **86**, 647, 1952.

[8]) C. Z. van Doorn and D. de Nobel, Physica **22**, 338-342, 1956 (No. 4).

[9]) F. A. Kröger, J. E. Hellingman and N. W. Smit, Physica **15**, 990-1018, 1949. N. W. Smit, Physica **16**, 317-328, 1950. F. A. Kröger and J. Dikhoff, Physica **16**, 297-316, 1950.

Nevertheless, a suspension of ZnS in a binder of this kind generally proves to have a fairly low breakdown voltage, because the ZnS grains with their conductive patches may easily form a bridge from one electrode to the other. Now the reflector layer of white lacquer (TiO_2 suspended in the same binder), which is sandwiched between the ZnS layer and the metallic electrode at the back, has a double function: besides reflecting the light emitted in the ZnS (towards the transparent electrode), it effects a considerable rise in the overall breakdown voltage. However, the reflector layer is not made unduly thick (not more than 20 μ), because of losses and the voltage drop across it which adversely affect the efficiency and emittance of the panel. Owing to the high dielectric constant of TiO_2, this drop is only about 25% of the overall voltage, but the reflector layer roughly doubles the breakdown voltage of the panel.

These layers are applied and baked-on successively. The whole is then backed with a metallic electrode by spraying with a colloidal suspension of silver. The panel is then thoroughly dried and, without delay, provided with a further backing in the shape of a glass plate, which is glued on to it; this is to prevent moisture from the atmosphere penetrating the somewhat hygroscopic layer of lacquer. Traces of moisture seriously diminish the life of the panel. If the layer is properly dried and protected, the life of the panel can extend to several thousands of hours, its emittance normally suffering a loss of about 10% within a short time (50 hours) of its being taken into use.

Good protection against the penetration of moisture can also be obtained by suspending the ZnS grains in a glaze instead of a lacquer; the suspension can be applied to a metallic base. However, in the present article we shall confine ourselves to electroluminescent layers having an organic binder.

The method of fabrication described above allows panels to be made in any desired shape and size, and this is of importance where the panel is to be used for instrument scale illumination, for example. It is equally easy to shape the panel into letters, figures and the like.

Emittance and efficiency

The flux Φ in lumens given by a luminous panel, is equal to the product of the emittance H and the area A of the panel. Another point of interest is its efficiency η, i.e. the ratio of flux Φ to the power P consumed (i.e. the number of lumens per watt).

The power dissipated by a lossy capacitor — the luminous panel is electrically equivalent to such a capacitor — can most easily be derived from the power factor $\tan\delta$, which is the ratio of the resistive component of current through the capacitor to the purely capacitive (wattless) current. With a voltage of r.m.s. value V across the panel, the power dissipated is:

$$P = V^2 \times 2\pi f C \tan\delta, \quad \ldots\ldots \quad (7)$$

and its efficiency is:

$$\eta = \frac{\Phi}{P} = \frac{HA}{V^2 \times 2\pi f C \tan\delta}, \quad \ldots \quad (8)$$

where f is the frequency of the alternating voltage and C is the capacitance of the panel.

Fig. 6 shows curves of H, $\tan\delta$ and η as functions of voltage, as derived from measurements on a certain electroluminescent layer. Curves showing how the same quantities vary with frequency appear in *fig. 7*.

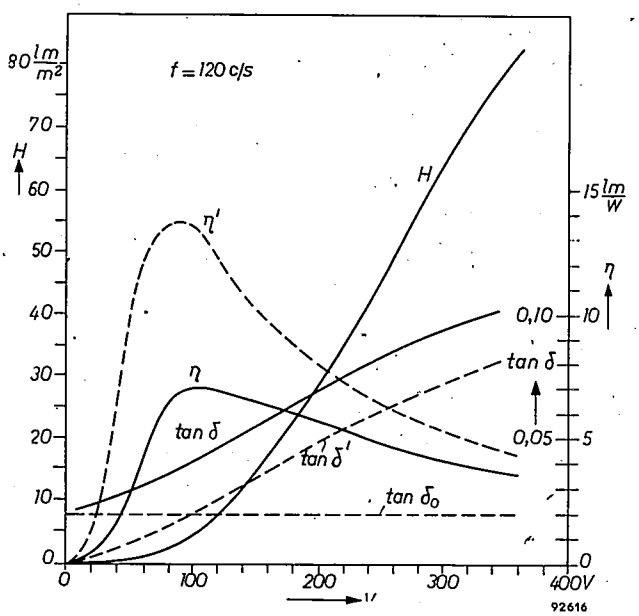


Fig. 6. Power factor $\tan\delta$, emittance H (in lumens/m²) and efficiency η (in lumens/watt) as functions of voltage V, the frequency being constant at $f = 120$ c/s, as obtained from measurements on a luminous panel giving green light. The dotted-line curves refer to the intrinsic efficiency η' (efficiency when the loss corresponding to $\tan\delta_0$ is neglected, see text).

Tan δ, the power factor, can be analyzed into two parts: a part $\tan\delta_0$, independent of voltage and proceeding from losses that are not directly connected with the mechanism of excitation (for the most part losses in the binder and in the series resistance of the tin oxide); and another part, $\tan\delta'$, which increases with voltage. This second part is connected with the power required for the transport of electrons through the ZnS crystals and for excitation. At higher voltages the resistive component of the

current corresponding to tan δ' increases more rapidly than the capacitive component (the rise of which with V is almost linear, since C is virtually independent of V); this indicates that the number of electrons injected increases with V at a more than linear rate. (Accordingly, we should, strictly speaking, introduce a correction into eq. (4) before comparing it with eq. (1); however, it can be shown that this correction does not essentially alter the voltage-dependence of the excitation as expressed by eq. (4) — see the article by Zalm cited in footnote [1]).)

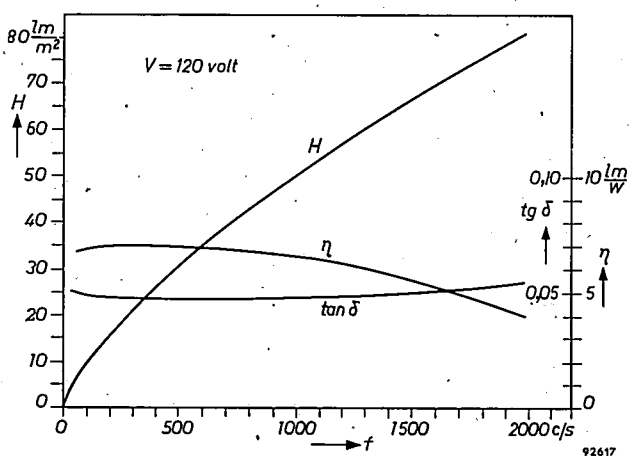


Fig. 7. The same quantities as in fig. 6 plotted as functions of f, the voltage being constant at $V = 120$ volts.

The maximum occurring in the curve of η as a function of V may be understood as follows. As a consequence of the way in which H depends on V (eq. 1), Φ increases rapidly with V at the start (at $V \approx 0$ it increases as an infinitely high power of V); as $V \to \infty$, however, the flux Φ approaches a limiting value, should electrical breakdown not already have taken place. Initially, the ratio $\Phi/P = \eta$ will likewise rapidly increase with V (as long as tan δ does not do so too), and it will attain a maximum at the point where numerator and denominator increase as the same power of V — that is, approximately at the point where Φ increases as the square of V. In reality, the maximum of η comes at a somewhat lower value of V, owing to the increase in tan δ with increasing V. For most electroluminescent layers, the quadratic increase of Φ with V occurs at a voltage between 100 and 200 volts.

As regards frequency dependence, it should be remarked that both C and tan δ are more or less independent of f. In electroluminescent layers with a large surface area, however, tan δ increases with frequency in consequence of the series resistance of the tin oxide layer. The measurements on which the present account is based were carried out on panels having an area of a few square centimetres only, in order to eliminate this complication. It will be clear why tan δ is independent of frequency if it be borne in mind that approximately equal charges are transported in each oscillatory period. Hence both the capacitive and resistive components of current increase more or less proportionally with frequency. Since Φ is approximately proportional to f, the efficiency η is more or less independent of f. The increase of Φ with f starts to become markedly less than linear at frequencies above 1000 c/s; the reason is that at these higher frequencies an ever smaller proportion of the electrons in the space-charge region has an opportunity to recombine with positive charges during the brief time that the field is of opposite sign. It will be seen that in oscillograms like that of fig. 2 the flashes begin to overlap considerably at these frequencies.

We may therefore conclude that maximum efficiency is obtained at moderate values of V, and hence for moderate values of emittance, and that at higher values of V higher values of H will be attained at the cost of lower efficiency. At higher frequencies H starts by increasing (linearly with f, roughly speaking) without η falling off. But in the kilocycle range η clearly begins to fall off; in panels with large surface areas the fall-off of η begins at still lower frequencies, because in such panels the series resistance of the tin oxide layer plays a bigger part.

Performance of existent electroluminescent layers

In recent years a considerable improvement in the emittance, efficiency and life of luminous panels has resulted from research [10]) into the optimum dimensioning of the panels as regards layer thickness, the filling fraction of ZnS in the binder, methods of applying the layers, and so on. In *Table I* some characteristic values are summarized for two panels having green emission (with and without an anti-breakdown TiO_2 reflector layer), and one panel having blue emission. The table shows the useful effect of the white TiO_2 layer in raising the breakdown voltage V_{cr}, the maximum efficiency η_m and the maximum emittance H_m attainable at low voltages. The blue panel has a much lower efficiency; this is due in part to the low sensitivity of the eye to blue light,

[10]) In the programme carried out in the Eindhoven laboratories, H. J. M. Joormann was responsible for the work on lacquers; measurements were carried out by T. J. Westerhof and K. W. C. Lugtenborg.

Table I. Data on ZnS luminous panels produced with urea-formaldehyde binder and having an area of about 7 cm². Each of the values given in the table is the average of those found for four panels made in the same way. In addition to the breakdown voltage V_{cr} and the relative dielectric constant ε, the table gives: the maximum efficiency (overall efficiency η_m, intrinsic efficiency η_m'), which is obtained at a voltage V_m; the emittance H_m obtained at the same voltage at a frequency of 50 c/s; emittance and efficiency at the maximum voltage permissible ($V = 0.6\ V_{cr}$) and a frequency of 50 c/s; and finally, emittance and efficiency at this maximum voltage and a frequency of 2000 c/s.

Composition of panel and thickness of layers	Colour of light emit.	V_{cr} (V)	ε	$V = V_m$ and $f = 50$ c/s			$V = 0.6\ V_{cr}$ and $f = 50$ c/s		$V = 0.6\ V_{cr}$ and $f = 2000$ c/s	
				η_m (lm/W)	η_m' (lm/W)	H_m (lm/m²)	H (lm/m²)	η (lm/W)	H (lm/m²)	η (lm/W)
45 μ ZnS	green	300	11	4.6	12	4	40	2	1000	1
29 μ ZnS + 23 μ TiO_2	green	550	12.8	10.2	17	4.6	70	4.5	1750	2
25 μ ZnS + 20 μ TiO_2	blue	400	9.8	0.8	—	0.3	3	—	90	—

in consequence of which fewer lumens are represented in a blue source than in a green one radiating the same amount of energy. Another reason for the low efficiency of the blue panel is that in the blue grains there is a higher power loss (higher tan δ). The η_m' column gives efficiency figures in the calculation of which only tan δ', the voltage-dependent part of the power factor, has been taken into account (for which, in other words, a fraction tan δ_0, caused by losses in the binder etc. having no connection with the mechanism of excitation, has been subtracted from the total power factor tan δ). It will be seen that even in the panel with the highest efficiency there is still a "useless" loss of 40%. If it were possible to eliminate this loss without reducing ε, the efficiency of the panel could be improved by a good 50%, thus becoming somewhat higher than that of an incandescent lamp. The efficiency of this hypothetical panel would approximate to the maximum efficiency theoretically obtainable from electroluminescence, as estimated by Zalm [1]).

The emittance of a panel connected to the 50 c/s mains, possibly via a transformer, can attain a value of 70 lm/m², i.e. about a tenth of the emittance of the screen of a television picture tube. Emittance can be greatly increased by connecting the panel to a generator of higher frequencies; supplied with A.C. at 2000 c/s, the best panel has an emittance of 1750 lm/m² and an efficiency roughly half of that which it has at 50 c/s (providing its surface area is not too large).

From these data it will be clear that for lighting a living room, for example, a panel area of the order of 1 square metre will certainly be necessary to obtain an adequate total luminous flux, and that the panel will have to be supplied at a frequency of some kilocycles, necessitating a special generator. The total efficiency of an installation like this, even for a panel emitting green light, will still be far below that of a normal incandescent lamp. This fact renders electroluminescence of little interest for normal lighting purposes. On the other hand, it offers an elegant solution to problems of scale illumination and the like: for this purpose there is no need for emittance and efficiency to be very high, the green colour is no objection, there is the advantage of easy shaping, and a further advantage deserving mention is the excellent uniformity of the brightness over the surface of a large panel.

However, the most interesting field of application of electroluminescent layers is undoubtedly their employment for intensifying images. To conclude this article, we shall give a short account of the "solid-state image intensifiers" based thereon. For the sake of brevity, and in order to distinguish these devices from image intensifiers of the vacuum-tube type [11]), we would propose for them the generic name of *amplificons* [12]).

Employment of electroluminescent layers for image intensification

There are two different methods of employing electroluminescent layers for this purpose. Both of them are, of course, based on a "control action" (as indeed are all forms of amplification); in other words, the energy flux is made to vary under the influence of a signal representing a very small, sometimes negligible energy. The first method can be referred to as that using external control, and the second as that using internal control.

---

[11]) M. C. Teves and T. Tol, Electronic intensification of fluoroscopic images, Philips tech. Rev. **14**, 33-43, 1952/53.

[12]) With regard to amplificons, see the article cited in footnote [2]) and also the December 1955 number of Proc. Inst. Rad. Engrs. (Solid-state materials issue).

Method using external control

Use is made of at least two layers, *I* and *II*, electrically connected in series (see *fig. 8*). Layer *II* is a normal electroluminescent layer having one electrode only, which is transparent. On top is applied a layer (*I*) of a photoconductive substance such as cadmium sulphide; *I* is also provided with one electrode which is transparent to the radiation (of intensity H_1) to be intensified. Voltages of V_1 and V_2 are set up across *I* and *II* respectively by applying a voltage $V = V_1 + V_2$ across the whole.

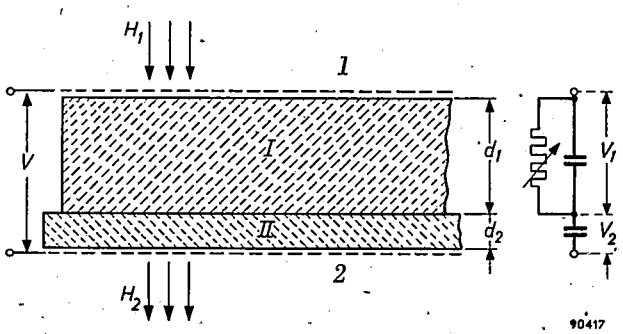


Fig. 8. Schematic diagram showing construction of an amplificon (solid-state image intensifier) using external control; the equivalent circuit is shown on the right. *1* and *2* transparent electrodes. *I* photoconductive layer of thickness d_1, across which acts a voltage V_1. *II* electroluminescent layer of thickness d_2, across which acts a voltage V_2. The two layers are electrically connected in series, and a voltage V ($= V_1 + V_2$) is applied across them. The radiation H_1 to be intensified falls on layer *I*; by virtue of electroluminescence in layer *II* an emittance H_2 is obtained.

In the absence of radiation a photoconductive substance possesses but few mobile electrons; when radiation containing quanta of sufficient energy falls upon it, however, some of its electrons are rendered mobile. Zinc sulphide too, to some extent, exhibits the phenomenon of photoconductivity (on this is based the functioning of the intensifier using internal control — see below), but cadmium sulphide does so to a far greater extent.

Now, the ratio between d_1 and d_2, the thicknesses of the two layers, can be made such that in the absence of incident radiation (that is, when $H_1 = 0$) almost the whole of the voltage applied acts across layer *I* (so that $V_2 \approx 0$ and $V_1 \approx V$). In these circumstances layer *II* emits little or no light (its emittance $H_2 \approx 0$). For this to be so, d_1 must be at least a few tenths of a millimetre thick for a d_2 of 50 microns.

If layer *I* is locally illuminated, the electrical resistance at the illuminated spots will diminish, and the greater the intensity H_1 of the incident radiation, the lower will be the resistance of the places affected. The result is that a smaller fraction of the applied voltage V acts across layer *I* at these places, V_2 becomes higher there and so does H_2. Thus layer *II* reproduces the local (and temporal) variations in H_1, and it does so in *intensified* measure: amplificons have already been made in which up to 50 quanta are emitted (H_2) for every quantum of incident radiation (H_1). The overall effect of the amplificon is described by the curve given by the relation between the local value of H_1 and the corresponding value of H_2. In *fig. 9* this curve has been constructed by combining the V_2-H_1 and H_2-V_2 characteristics, the latter being the "electroluminescence characteristic".

It will be observed that the relation between H_1 and H_2 is not linear, and this means that generally the gradation in the incident image is not faithfully reproduced. In principle, this makes it possible to heighten contrast, and the capacity to do so can be of importance in the intensification of X-ray images. However, by adopting special measures, such as the use of an alternating voltage containing two or more frequencies, the characteristic can be considerably straightened out.

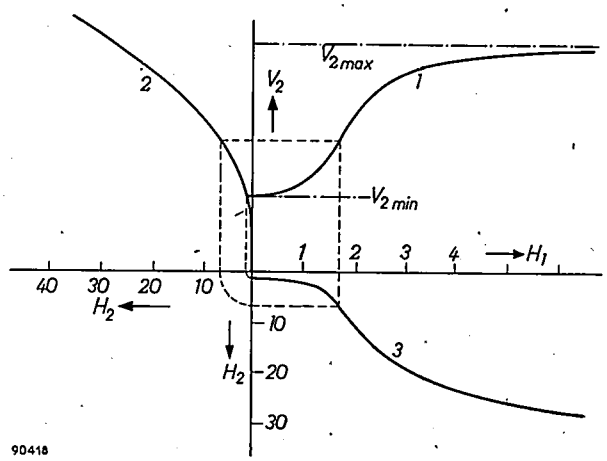


Fig. 9. Amplificon characteristic (*3*) constructed by combining the V_2-H_1 characteristic (*1*) and the H_2-V_2 or electroluminescence characteristic (*2*). $V_{2\,max}$ = maximum voltage acting across layer *II* when the radiation falling on layer *I* is very intense (when $H_1 \rightarrow \infty$). $V_{2\,min}$ = minimum voltage acting across layer *II* in the absence of incident radiation (when $H_1 = 0$).

Method using internal control

Destriau [13]) discovered that certain zinc sulphides activated with copper and manganese, which exhibit but little electroluminescence even at high voltages, might be caused by irradiation with ultra-violet or X-rays to emit light more strongly. In this case too, electroluminescence reproduces local and temporal variations in the

[13]) See p. 1911 of the article by Destriau and Ivey referred to in footnote [1]).

incident radiation, giving "intensification" of this radiation. The functioning of this kind of intensifier is to be understood as follows. The ultra-violet quanta absorbed by the intensifier render additional electrons mobile in the ZnS, thus furthering the mechanism described above by which space-charge layers are formed. In this way, the controlling of the process of electroluminescence takes place within the ZnS itself. In an image intensifier constructed on this principle by Cusano [14]), up to 10 visible quanta are emitted by the panel for every quantum of ultra-violet radiation it absorbs. At high levels of illumination the intensification effect falls off. The presence of manganese in the layer allows the intensifier to operate on D.C. as well as A.C.

For a more detailed discussion of amplificon design and performance the reader is referred to the literature as indicated in the footnotes. Although these types of image intensifier are still in their infancy and it is impossible to say with certainty how long it will take for them to evolve into a product suitable for manufacture, it is nevertheless very probable that they will find many applications in the future. It is conceivable that radar reception, fluoroscopy and perhaps television will be amongst them. For the last-mentioned, the inertia of the effect (namely, of the photoconduction process in the intensifier using external control), which still constitutes an obstacle at the present time, would have to be overcome.

Summary. Layers of specially-prepared zinc sulphide powder, activated with copper, emit light when an electric alternating tension is applied to them. This effect, which is called electroluminescence and which was discovered by Destriau in 1936, can be applied to dial illumination in instruments, radio receivers and clocks and in general wherever uniformly luminous surfaces are required and where high efficiency and high luminous flux per unit area (emittance) are not essential. This method of light production does not promise much as far as general lighting purposes are concerned: at best (in layers emitting green light) its efficiency is just about equal to that of the incandescent lamp and, in order to raise emittance to a value high enough for the adequate lighting of a living room (for example) with a luminous surface of not more than, say, a square metre in area, the electrical supply has to be at a fairly high frequency, e.g. 2000 c/s, and the efficiency is then considerably reduced. In this article the more important details of the complicated phenomenon of electroluminescence, such as the way emittance and efficiency depend on voltage and frequency, are explained in terms of a mechanism whose validity is supported by a number of experimental facts. The essential feature of this mechanism is the formation, near the surface of each zinc-sulphide grain, of thin layers of strong positive space charge, where the field strength is much higher than the average; by virtue of the strong field, mobile electrons in the layers are capable of bringing about intense excitation by impact and, each time the field reverses, a certain proportion of the bound positive charges in the layers emit light on recombining with electrons. Finally, a brief account is given of the employment of the phenomenon in solid-state image intensifiers (amplificons), which may well find important applications in radar, X-ray fluoroscopy, etc.

[14]) D. A. Cusano, Phys. Rev. **98**, 546-547, 1955.

THE SUB-MICROSCOPIC STRUCTURE OF "TICONAL" G MAGNET STEEL

620.18:669.15.255.24.71.018.58:621.318.2

"Ticonal" G is an alloy for permanent magnets, composed of 51% Fe, 24% Co, 14% Ni, 8% Al and 3% Cu [1]). When this alloy is cooled at a predetermined rate in a magnetic field and then annealed, an anisotropic material with excellent magnetic properties is obtained [2]): the value of $(BH)_{max}$ for this material amounts to about 5.0×10^6 gauss.oersted, measured in the direction of the field during cooling. Even better results can be reached with single crystals of this material (which crystallizes in the cubic structure) if it is ensured that during cooling down a (100)-direction is parallel to the direction of the field [3]). The $(BH)_{max}$ can thus attain a value of 8.0×10^6 gauss.oersted.

Fig. 1 is a photo-montage giving a three-dimensional impression of the metallographic structure of a single crystal of this kind. It has been obtained by placing electron-microscope photographs of the structure of the (100)-planes on the faces of a wooden cube, the direction of the magnetic field during cooling being indicated by the arrow. The photographs concerned are shown separately in *figs. 2* and *3*.

They reveal a substantial difference between the structures of a (100)-plane parallel to and perpendic-

[1]) The alloy known in British and American literature as Alnico V is an identical alloy.

[2]) B. Jonas and H. J. Meerkamp van Embden, Philips tech. Rev. **6**, 8-11, 1941.

[3]) Netherlands patent No. 71925; see Philips tech. Rev. **18**, 358-360, 1956/57 (No. 12).

ular to the magnetic field. The former is characterized by elongated elements with the longest dimension in the direction of the magnetic field (fig. 2). The (100)-planes perpendicular to the magnetic field during cooling are characterized by a structure in which mutually perpendicular elements predominate which, however, exhibit no pronounced elongated form (fig. 3).

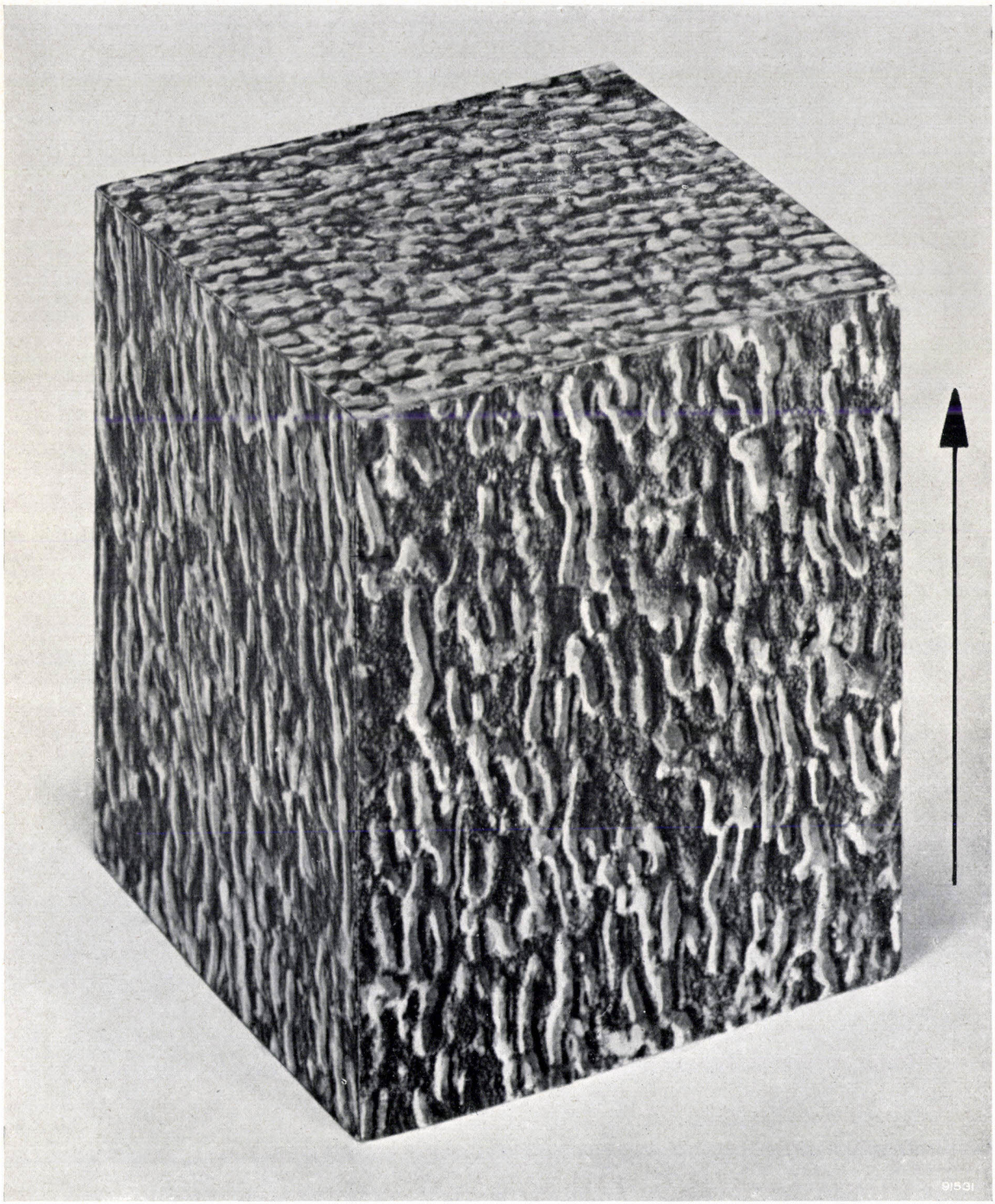


Fig. 1. Photo-montage giving a three-dimensional impression of the metallographic structure of a single crystal of "Ticonal" G. The arrow indicates the direction of the magnetic field applied during the cooling period. This picture was obtained by placing the structural photographs (magnification 160 000 diameters) shown in figs. 2 and 3 on the corresponding faces of a wooden cube.

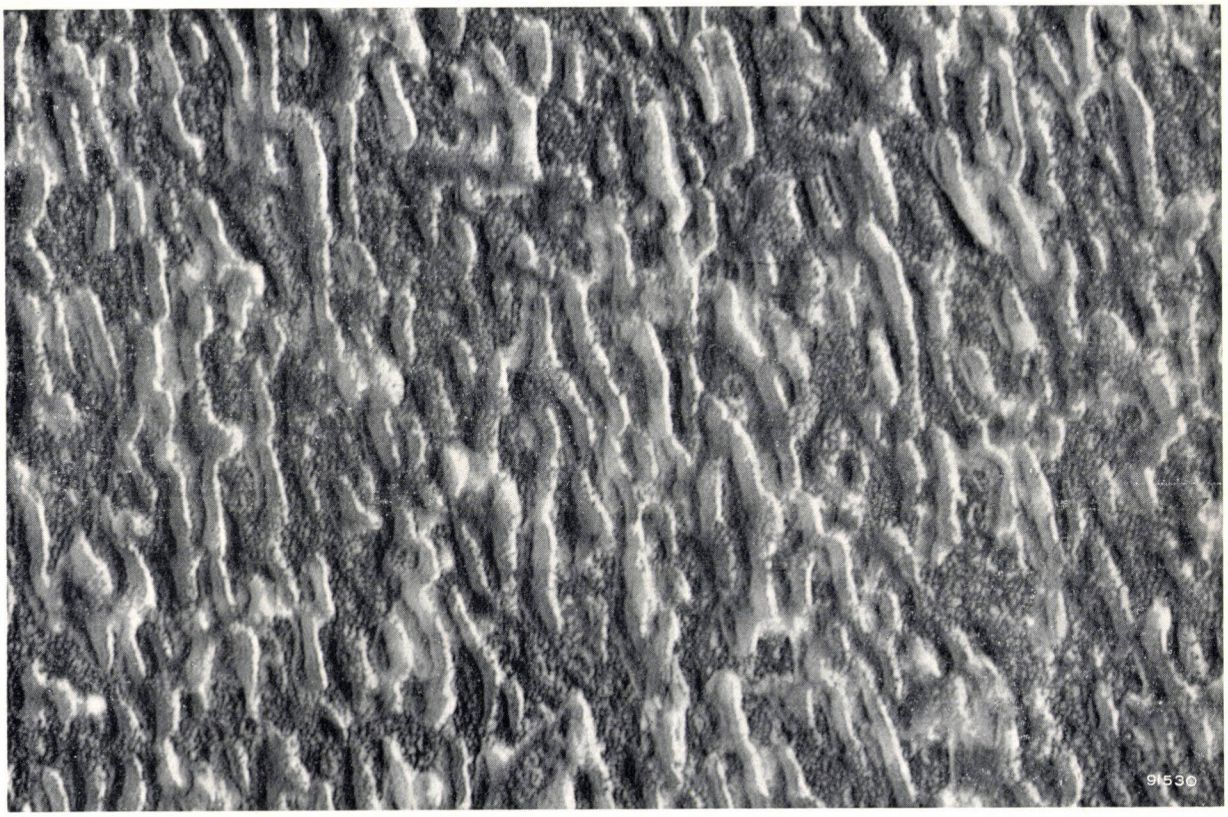


Fig. 2. Electron-microscope photograph of the structure of a (100)-plane *parallel* to the external magnetic field applied during cooling. Direct carbon replica of a platinum-shadowed, electrolytically polished surface, etched with a dilute solution of nitric acid in alcohol. Magnification 160 000×. This photograph was made with the Philips electron microscope EM 100 b.

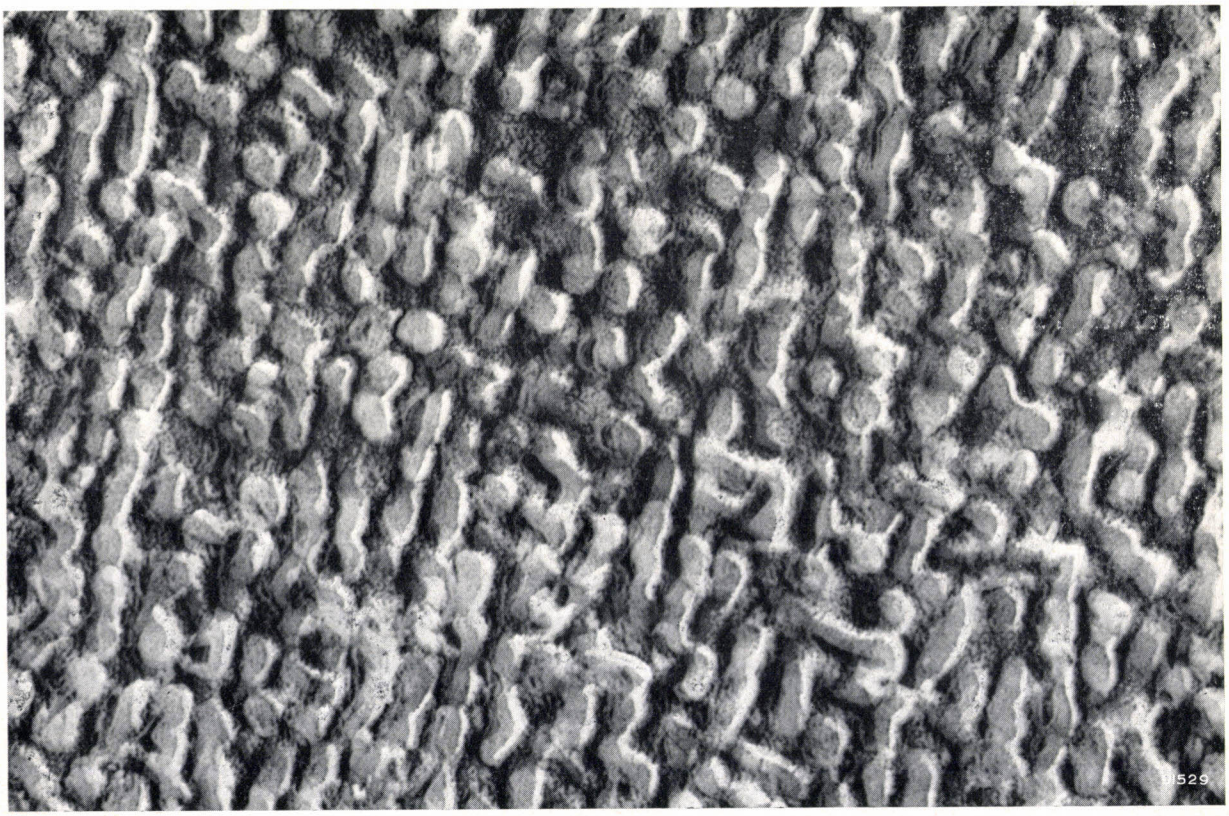


Fig. 3. As fig. 2, but for a cube face *perpendicular* to the magnetic field.

In describing the structure, which indeed corresponds to the earlier formed concept based on the magnetic behaviour of this material, the term "structural elements" has been used deliberately, leaving open the question whether a well-defined second phase has been precipitated, or whether we are concerned with a pre-precipitate stage characterized merely by periodic fluctuations in the composition. The structural pictures obtained do not indicate which of these two ideas is correct, since even with a gradually changing composition sudden changes in the susceptibility to chemical attack may occur, such that etching produces discontinuities which resemble phase boundaries (Tammann's boundaries of chemical resistivity).

The extremely fine structural details between the relatively coarser "grains" have not yet been satisfactorily explained.

It is not possible within the space of this short report to compare the present results with those obtained by other workers [4]). Another paper, now in preparation, will go further into these matters.

H. B. HAANSTRA, J. J. de JONG and J. M. G. SMEETS.

[4]) R. D. Heidenreich and E. A. Nesbitt, J. appl. Phys. **23**, 352-365, 1952.
K. J. Kronenberg, Z. Metallkunde **45**, 440-447, 1954.
H. Fahlenbrach, Tech. Mitt. Krupp **12**, 177-184, 1954; **14**, 12-15, 1956 (No. 1).
D. Schulze, Exp. Techn. Physik **4**, 193-204, 1956 (No. 5).

THE JUNCTION TRANSISTOR AS A NETWORK ELEMENT AT LOW FREQUENCIES

I. CHARACTERISTICS AND *h* PARAMETERS

by J. P. BEIJERSBERGEN, M. BEUN and J. te WINKEL. 621.375.4

Transistors of various types are now becoming available in large numbers. In Europe, the Philips semiconductor factory at Nijmegen (to name one centre), which started up on the 12th July 1955, has been working to capacity for some time. Equipment embodying transistors has on several occasions been dealt with in this periodical. It seemed to the editors desirable to give an outline of the elementary properties of the transistor as a network element, in order to provide a basis for later articles on transistor applications. The article which follows, and which falls within the framework of the series of transistor articles started in 1955, is the first part of this outline.

Transistor characteristics

The physical basis of the functioning of the junction transistor at low frequencies has been discussed at length in two articles published earlier in this Review [1] [2]). We shall now turn to the characteristics of the transistor at low frequencies, leaving the physical mechanism of the device out of consideration. We look at the transistor as a network element having three connecting wires, marked by the manufacturer as the emitter, base and collector connections. Hence three different currents and three different voltages are involved, a total of six variables (*fig. 1*). If we apply direct voltages between collector and emitter (V_{ce}) and between base and emitter (V_{be}) (by so doing we fix the third voltage V_{cb} between collector and base, for, in accordance with Kirchhoff, $V_{ce} = V_{cb} + V_{be}$), the three currents will have values determined by the voltages [3]). As regards the three currents, we know from Kirchhoff's first law that $I_c + I_b + I_e = 0$. If we plot two of them, e.g. I_b and I_c, as functions of V_{ce}, for various fixed values of V_{be}, we shall have the properties of the transistor in the form of two families of (static) characteristic curves (*fig. 2*).

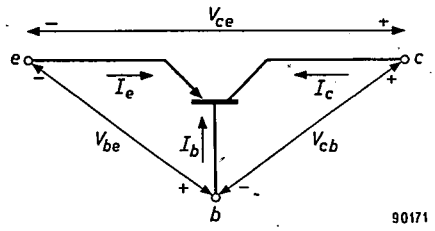


Fig. 1. The three voltages and three currents occurring as variables in transistors. *e* is the emitter, *c* the collector and *b* the base terminal. In accordance with the usual convention, the emitter terminal has an arrowhead to distinguish it from the collector terminal. The direction of the arrow shows whether the transistor is a *P-N-P* or *N-P-N* type; in the former case it points towards the base, in the latter case away from it.

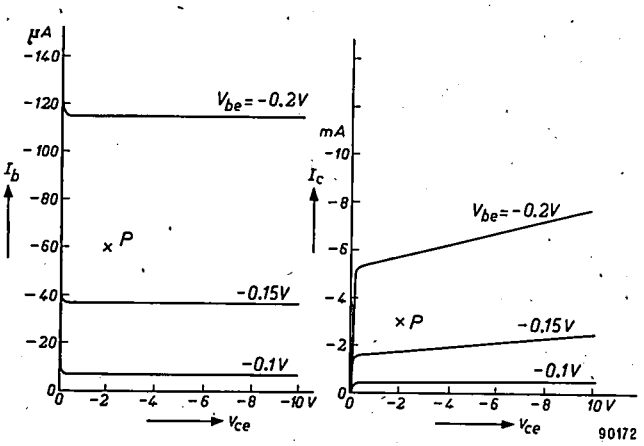


Fig. 2. Example of a pair of families of curves completely describing the properties of a transistor — here the Philips OC 71 — at low frequencies. The OC 71 being a transistor of the *P-N-P* type, all the quantities in the above graphs turn out to be negative. The biassing of the transistor might give an operating point at *P*, for example.

It is hardly necessary to point out that, when applied to A.C. signals, these static curves provide a valid approximation to the transistor properties only up to a certain limited frequency. Where that fre-

[1]) F. H. Stieltjes and L. J. Tummers, Simple theory of the junction transistor, Philips tech. Rev. **17**, 233-246, 1955/56.

[2]) F. H. Stieltjes and L. J. Tummers, Behaviour of the transistor at high current densities, Philips tech. Rev. **18**, 61-68, 1956/57 (No. 2).

[3]) The sequence of the indices (b for base, c for collector and e for emitter) implies a convention as to the sign of the voltages: for example, a positive value of V_{ce} signifies that the potential of the collector is higher than that of the emitter. Thus, for example, $V_{ce} = -V_{ec}$. Currents are regarded as positive when they flow towards the transistor — see the small arrows in fig. 1.

quency limit lies depends very largely on the type of transistor. For the OC 71, which we shall continually be taking as an example in this article, the limit may be placed at a frequency of 10 kc/s; for transistors that are specially designed for high frequencies the limit is very much higher.

The OC 71 is a transistor of the *P-N-P* type (see article [1])); accordingly, the transistor symbol in fig. 1 has the arrowhead that distinguishes the emitter from the collector pointing inwards. This is a convention in general use. Between *P-N-P* and *N-P-N* transistors there are no essential differences with regard to their behaviour as network elements.

Fig. 2 represents only two of the many possible families of curves. It is found, in fact, that 54 different families of curves can be drawn for one transistor [4]). Two mutually independent families of curves (i.e. such that one of the pair cannot be derived from the other) are sufficient to describe the transistor's properties at low frequencies. Which families of curves will be most appropriate and useful depends on circumstances. At this stage, suffice it to say that the pair shown in fig. 2 will seldom be encountered; the reason for their choice on this occasion is that the family of curves on the right corresponds to the usual anode characteristics of a tube, as will become clear in the next section.

In use, a transistor — like a tube — is biassed with certain direct currents and voltages. In any family of curves a point can be found that corresponds to these values — the operating point. It is of importance to observe that, when the transistor is biassed to a normal D.C. operating point (e.g. to *P* in fig. 2), the base current is always small in comparison with both the emitter current and the collector current (see the numerical values in fig. 2; for the physical reasons refer to article [1])). Hence the current that flows into the transistor at the emitter leaves at the collector almost unchanged in value. As to the voltages, in germanium transistors the voltage V_{be} between base and emitter is low in comparison with the voltages which those two electrodes have with respect to the collector. Thus $V_{bc} \approx V_{ec}$.

[4]) A. J. W. M. van Overbeek, Enige schakelingen met transistors, T. Ned. Radiogenootschap **19**, 231-260, 1954. We have $n = 6$ variables (the transistor voltages and currents) and require to know the number of ways in which any three of these may be selected and arranged between $m = 3$ locations (the abscissa, ordinate and parameter). This is given by $n!/(n-m)! = 120$ ways. Of these, the $3! = 6$ arrangements of the three voltages alone, and likewise the six arrangements of the three currents alone, merely give expression to Kirchhoff's laws and are therefore trivial, for they give no information about the transistor itself. There remain 108 graphs with meaningful content. However, this number is made up of pairs which are identical except that the ordinate of one graph is the abscissa of the other and *vice versa*.

A comparison with tubes

A triode, like a transistor, is a network element with three terminals; again, therefore, its properties can be expressed on paper in the form of two families of curves, the curves, for example, of grid current I_g and anode current I_a as functions of anode voltage V_{ak} (i.e. anode-cathode voltage), with the grid voltage V_{gk} (i.e. grid-cathode voltage) as the running parameter. The latter family constitutes the usual anode characteristics of a tube. A pentode has five terminals, and in theory there are many more current and voltage variables involved than in a triode or transistor. However, elementary treatment of the pentode is usually confined to the case in which the screen and suppressor grids have fixed potentials with respect to the cathode; in this way we arrive at the network element shown in *fig. 3*, and this again is one with three terminals. The properties of such a pentode can likewise be expressed in the form of two families of curves.

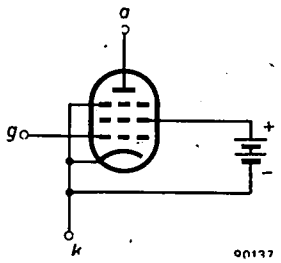


Fig. 3. A pentode has more than three terminals, but if screen and suppressor grids are given fixed potentials with respect to the cathode only three terminals are left, as is the case with a triode or transistor. Through the terminal marked *k* flows the sum of the cathode, screen-grid and suppressor-grid currents.

In certain respects transistors and tubes exhibit a considerable similarity in properties. The grid current of tubes is very small and hence the other two currents are roughly equal. Again, the voltage between grid and cathode is very low compared with the voltages between those electrodes and the anode, so that $V_{ga} \approx V_{ka}$. Thus, in comparing a transistor with a tube, the best analogy is obtained and the differences are brought out most clearly by regarding the base as corresponding to the grid, the emitter to the cathode, and the collector to the anode. Thus the curves plotted on the right of fig. 2 correspond to the anode characteristics of tubes.

One of the striking differences exhibited by tubes is that, in cases where the grid voltage is negative with respect to the cathode, the grid current is not merely small — it is quite negligible (the base current in a transistor, though small, is not negligible). Therefore, if we are only interested in applications in which the grid voltage remains negative (and this is so in receivers and amplifiers, but not usually in

transmitters), we have one variable less, and only one family of curves is required for expressing the properties of the tube. This is one of the reasons why it is often easier to understand the behaviour of tubes than that of transistors.

The common-emitter, common-base and common-collector configurations

We have seen that relations between only four of the six variables (two being voltages and two currents, as in fig. 2) suffice to describe a transistor completely. There are three simple cases where it is immediately clear which four variables it will be best to select; these cases arise when the transistor is inserted between two 2-terminal networks, in the manner shown in *fig. 4a*. Here the transistor itself is regarded as a 4-terminal network, or fourpole, that is to say a network in which two terminal pairs can be distinguished [5]). Like any other network element having three terminals, a transistor can be made into a fourpole in three different ways (figs. 4*b*, *c* and *d*). In fig. 4*b* the emitter has been given two terminals. One of them, together with the base, constitutes the first terminal pair; the other, together with the collector terminal, constitutes the second pair. Since the emitter connection is shared between the two pairs of terminals, the transistor is said to be in the common-emitter (also "*grounded* emitter") configuration. Alternatively, double terminals can be given to the base or to the collector, and doing so produces the common-base (fig. 4*c*) and common-collector (fig. 4*d*) configurations respectively.

In the common-emitter configuration, V_{be} and I_b are the voltage between and current through the one pair of terminals, and V_{ce} and I_c the voltage between and current through the other pair. Clearly, for this configuration we shall want to make use of those relations (i.e. those families of curves) that relate the voltages and currents just mentioned. It may easily be found from figs. 4*c* and 4*d* which currents and voltages should occur in the curves appropriate to the two other configurations.

The three simple circuits just described provide a suitable basis for discussing some of the fundamental properties of the transistor as a network element. In all three configurations a transistor can act as a

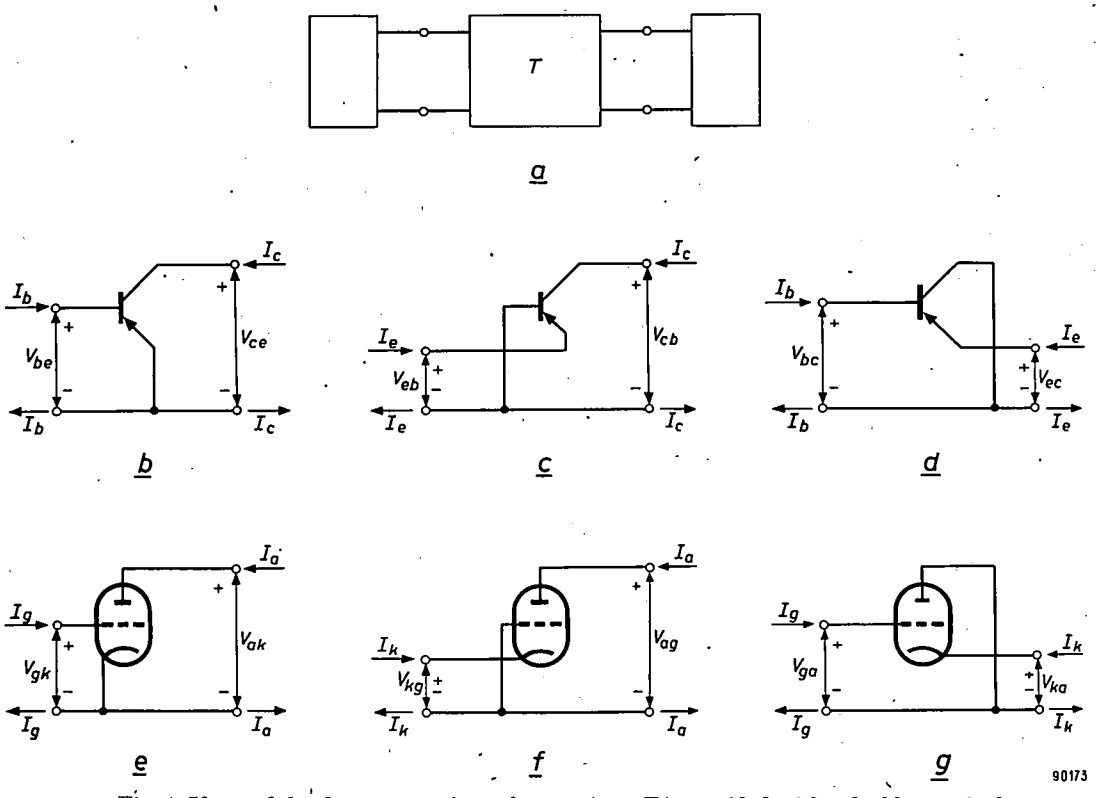


Fig. 4. If one of the three connections of a transistor T is provided with a double terminal, the transistor can be treated as a fourpole and inserted between two twopoles, as in (*a*). Three "configurations" are produced, (*b*) the common-emitter configuration (i.e. the emitter is common to input and output), (*c*) the common-base configuration and (*d*) the common-collector configuration, according to whether emitter, base or collector is given the double terminal. The corresponding tube circuits are also given — (*e*) common cathode (also "grounded" cathode), (*f*) common grid and (*g*) common anode (cathode follower).

[5]) Two terminals constitute a terminal pair when the current that enters by one of them leaves via the other.

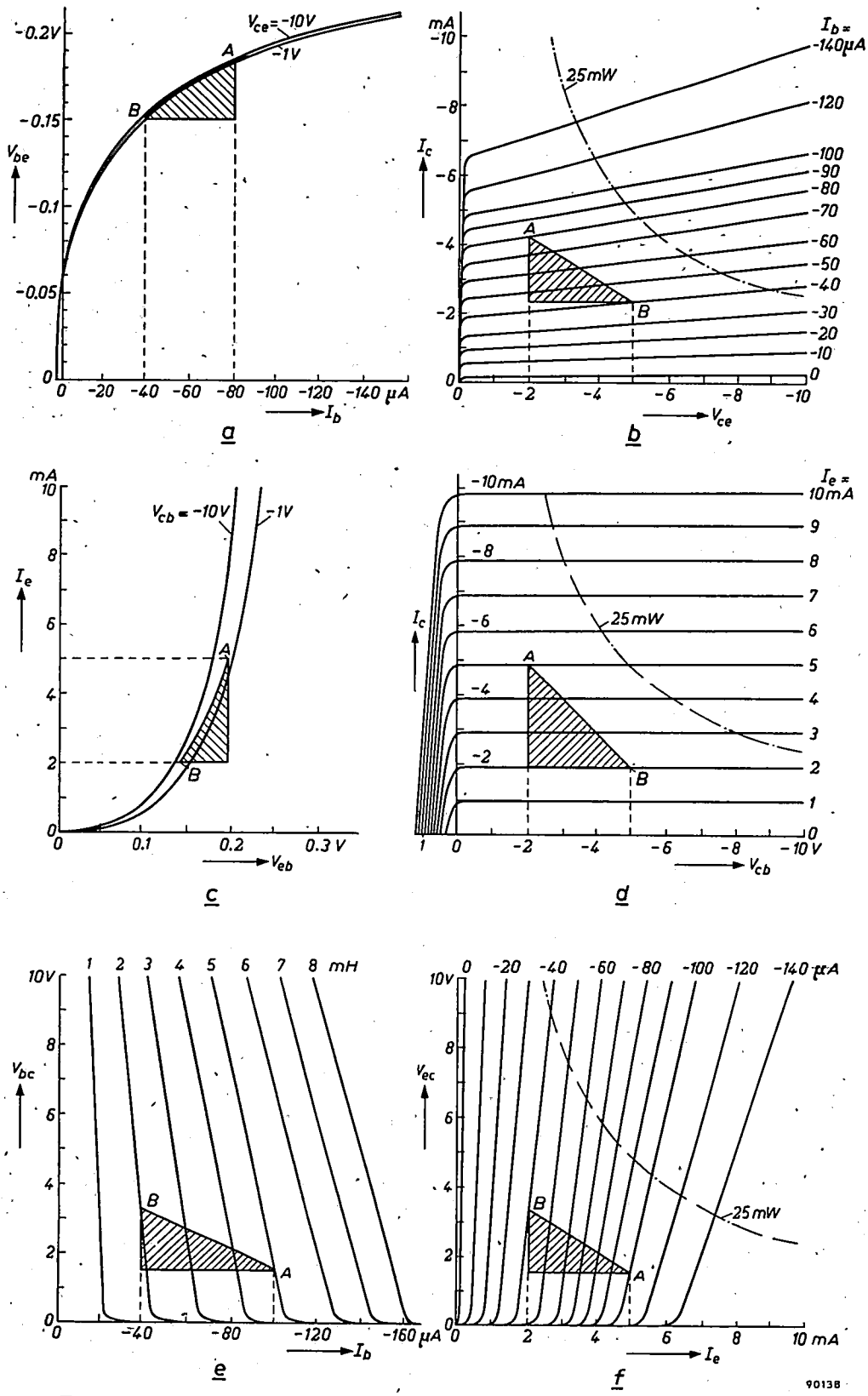


Fig. 5. Input characteristics (left) and output characteristics (right) of the OC 71 transistor in (*a*) and (*b*) the common-emitter, (*c*) and (*d*) the common-base, and (*e*) and (*f*) the common-collector configurations. These curves have been drawn to show that a power gain is obtainable in each of the three configurations. The load is imagined to be a pure resistance. *A* and *B* represent the points between which the instantaneous values of current and voltage vary as a result of the superposition of alternating current on the D.C. bias operating point. The areas of the hatched triangles on the left represent power taken up at the input; the areas of those on the right represent power delivered at the output. Differences in scale values should be noted. The hyperbolae in the output graphs are lines of constant power dissipation.

power amplifier. In the common-emitter configuration (fig. 4*b*), the direct voltage between and current through the base-emitter terminal pair are low in comparison with the direct voltage between and current through the other pair. This suggests that the same will be true of alternating currents and voltages superimposed on the D.C. operating point values. When base and emitter form the input pair and collector and emitter the output pair, therefore, we may expect power amplification to take place by way of simultaneous voltage and current amplification. This is confirmed by a glance at *figs. 5a* and *b*, representing the input and output characteristics for the configuration in question. If it is assumed that output voltage and current vary between the values represented by points A and B in fig. 5*b* and if the corresponding points A and B in fig. 5*a* are worked out, it is immediately clear that small variations at the input give rise to much larger variations at the output.

It will be of value here to put the corresponding tube circuit alongside the transistor configuration. In accordance with the preceding section, the corresponding tube circuit is that in which the cathode is the common electrode (this being the normal arrangement for a tube — see fig. 4*e* — viz. the common-cathode circuit or, more often, grounded-cathode circuit). In this the grid and cathode become the input terminals (small current and low voltage), and the anode and cathode become the output terminals (large current and high voltage).

In the common-base configuration (fig. 4*c*) the direct *currents* through the two terminal pairs (i.e. the emitter and collector currents) are of about the same magnitude; but the direct *voltage* between collector and base is much higher than that between emitter and base. That the same applies to alternating currents may be deduced from figs. 5*c* and *d*, by proceeding in the same way as above for figs. 5*a* and *b*. Thus, by choosing emitter and base as input pair and collector and base as output pair, power amplification is obtained as a consequence of voltage amplification. The tube circuit corresponding to this is the "grounded" grid (common-grid) circuit (fig. 4*f*). Here too, input and output currents have roughly the same magnitude, but the output voltage is much higher than the input voltage.

Finally we come to the common-collector configuration (fig. 4*d*). In this the voltages between the two pairs of terminals are roughly equal, but the currents differ. Power amplification consequent on *current* amplification is to be expected from this configuration. Base and collector are taken as the input terminals (the small base current becoming the input current), and emitter and collector as output terminals (the much larger emitter current becoming the output current). Figs. 5*e* and *f* confirm that a current amplification is obtained. The corresponding tube circuit is the common-anode circuit (better known as cathode-follower circuit — see fig. 4*g*), in which it is easy to see points of resemblance.

The terms "common-emitter", "common-base" and "common-collector" are frequently applied to circuits more complex than that of fig. 4*a*. Thus, the circuit shown in *fig. 6*, one that arises very frequently

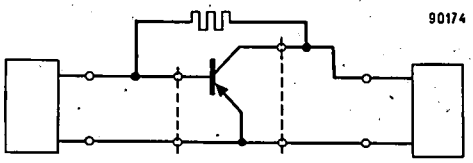


Fig. 6. A circuit frequently arising in practice, in which a (large) resistor is connected between collector and base. It is not possible here, without employing an artifice, to distinguish two pairs of terminals in the sense of footnote [5]).

in practice, is also referred to as a common-emitter arrangement; but it includes a resistor for negative feedback between collector and base, and it is not possible (at least, not without employing an artifice) to distinguish pairs of terminals in the sense of footnote [5]). In fact, the term "common emitter" is used in a wider sense to indicate that, in analyzing the circuit, use will be made of the currents and voltages associated with that configuration, and of the relations between those currents and voltages. The fourpole between two twopoles in fig. 6 naturally has characteristics different from those of the transistor standing alone. Changes in characteristics also result from modifications to the common-base and common-collector configurations and, *mutatis mutandis*, when modifications are made to the common-cathode, common-grid or common-anode circuits of tubes.

In what follows we shall mainly be concerned with the common-emitter configuration because this is the most commonly employed, albeit often in modified form (as in fig. 6, for example). In discussing this configuration we shall confine ourselves to variations of voltage and current so small that the transistor can be regarded as a linear network element.

The possible number of families of characteristics connecting the four interrelated currents and voltages in the common-emitter configuration is 12 (being $\frac{1}{2} \times 4!/(4-3)!$ — see footnote [4])). Most of them are used only incidentally, according to need and to taste, in order to investigate or explain the properties of transistors. It will be seen that the output characteristics in fig. 5*b* (where, in contrast to the usual practice with tubes, the input *current*, not the input *voltage*, is the

running parameter of the curve family) are remarkably straight and equidistant throughout a considerable range of currents and voltages. This signifies that within this range, provided that it is permissible to regard the input *current* as the input signal carrier (i.e. to regard the transistor as being "current driven"), the transistor introduces but little distortion. Against this, the input characteristics (fig. 5*a*) have considerable curvature; they are in fact exponential functions, as may be deduced from the physics of the transistor (see [1])). Hence, unless the amplitude of the input signal is very small, a sinusoidal input *current* implies a non-sinusoidal input *voltage*. Insofar as the input signal is an input *voltage*, therefore (i.e. when the transistor is "voltage driven"), considerable distortion may arise. For this reason it may be said that the transistor is first and foremost a current amplifier. In order to get an idea of the situation when the transistor is voltage driven — which in tubes is practically the only case arising — it will naturally be better to draw output characteristics with the input voltage as parameter, and these will give a less favourable impression of the linearity of a transistor. Even the characteristics of fig. 5*b*, however, do not remain equidistant at high values of I_b, but bunch together ever more closely. It is even possible to have non-linearity with respect to input *voltage* and non-linearity with respect to input *current* compensating each other to a considerable extent. The question of distortion in the cases where control of the transistor is neither by current alone nor by voltage alone, is a complex problem, the full treatment of which lies outside the scope of this article.

However, the reason why the input current is almost always taken as the running parameter in transistor output characteristics is connected not so much with distortion as with the way in which the transistor is biassed to the desired operating point. In choosing the operating point, use is made of the output graph (fig. 5*b*) because lines of constant power dissipation, i.e. hyperbolae of the form $I_c \times V_{ce} =$ constant, can easily be drawn in it. (The total power dissipated in the transistor actually amounts to $I_b \times V_{be} + I_c \times V_{ce}$, but the first term can be neglected.) In this way it can be seen immediately whether any operating point that may be chosen will involve excessive power dissipation. An impression is also gained of the extent to which current and voltage can be allowed to vary about the operating point without serious distortion arising. *Fig. 7* shows one possible method of biassing the transistor to the selected values of I_c and V_{ce}. The battery voltage V_0 (6 V, say) is high in comparison with V_{be}, the voltage between base and emitter (0.1 to 0.2 V), and hence $R_b \approx V_0/I_b$. To obtain a numerical value for R_b, one must be able to read off I_b from the graph; that is to say, I_b should be taken as the running parameter.

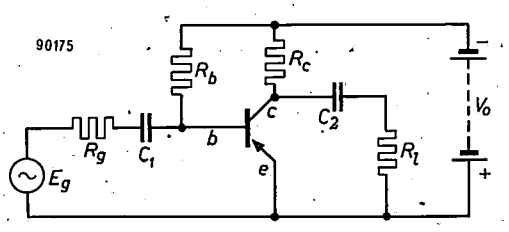


Fig. 7. A circuit with a transistor in the common-emitter configuration and with provision for the biassing. E_g is the e.m.f. and R_g the internal resistance of the signal source, R_l is the load, C_1 and C_2 are blocking capacitors, and R_b and R_c are resistors for setting the biassing. V_0: battery voltage.

Transistor fourpole parameters

If input voltage and current are denoted by V_1 and I_1, and output voltage and current by V_2 and I_2, the families of characteristics in figs 5*a* and *b* are graphical representations (for the transistor in the common-emitter configuration) of the relations:

$$V_1 = f(I_1, V_2), \quad \ldots\ldots \quad (1a)$$

$$I_2 = g(I_1, V_2). \quad \ldots\ldots \quad (1b)$$

It will become clear later on (page 22, column 1) why we have chosen I_1 and V_2 as the independent variables and V_1 and I_2 as the dependent variables. If current and voltage variations about the operating point are small enough for the relation between them to be regarded as linear, then we may write:

$$\Delta V_1 = \left(\frac{\partial V_1}{\partial I_1}\right)_{V_2} \Delta I_1 + \left(\frac{\partial V_1}{\partial V_2}\right)_{I_1} \Delta V_2,$$

$$\Delta I_2 = \left(\frac{\partial I_2}{\partial I_1}\right)_{V_2} \Delta I_1 + \left(\frac{\partial I_2}{\partial V_2}\right)_{I_1} \Delta V_2.$$

Introducing symbols whose definitions will be clear without further explanation, we can rewrite the above as follows:

$$v_1 = h_{11} i_1 + h_{12} v_2, \quad \ldots\ldots \quad (2a)$$

$$i_2 = h_{21} i_1 + h_{22} v_2. \quad \ldots\ldots \quad (2b)$$

The h parameters, as they are termed, occurring in these two equations, have meanings that can be deduced from the equations themselves. If v_2 is made zero, that is, if V_2 is kept constant (this amounts to short-circuiting the output for alternating current), it will be seen that:

$h_{11} = \left(\frac{v_1}{i_1}\right)_{v_2=0} =$ Input resistance with output short-circuited.

$h_{21} = \left(\frac{i_2}{i_1}\right)_{v_2=0} =$ Current amplification with output short-circuited; in these circumstances the usual term employed is "current amplification *factor*".

If i_1 is made zero, that is, if I_1 is kept constant — and this amounts to open-circuiting the input for A.C. — we see that:

$h_{12} = \left(\frac{v_1}{v_2}\right)_{i_1=0} =$ Feedback effect of output voltage on input voltage (reverse voltage amplification factor) with input open-circuited.

$h_{22} = \left(\frac{i_2}{v_2}\right)_{i_1=0} =$ Output conductance (reciprocal of output resistance) with input open-circuited.

h parameters can of course be introduced for any fourpole, including each of the three possible configurations of a transistor. In what follows we shall, where necessary, indicate common-emitter h parameters by adding an index "e" to the h; thus, e.g., h^e. (Where the common-base or common-collector configurations are involved, we can write h^b and h^c [6].) The h parameters owe their name to the fact that a current and a voltage, I_1 and V_2, occur in them as independent variables, this giving them a certain *h*ybrid character.

Relations (2a) and (2b) remain valid when i_1, i_2, v_1 and v_2 are functions of time, providing they change so slowly — i.e. provided their frequencies are so low — that the relations (1a) and (1b) as given by the static characteristics continue to be obeyed at all times. In this article we shall only consider such low frequencies. It is quite possible to write relations of the form of (2) at higher frequencies, but then the h parameters are complex numbers with a complicated relationship to frequency. At low frequencies the h parameters are real numbers, as is apparent from their significance as the slopes of curves.

We started our reasoning with relations (1a) and (1b), taking I_1 and V_2 as independent variables, because we wanted to arrive at the h parameters. There are six different ways of choosing a pair of independent variables from amongst I_1, I_2, V_1 and V_2, and hence six sets of parameters are obtainable for the same fourpole. As always, it depends on the problem to be solved which choice will make calculation simplest. In the design of a simple amplifier equipped with transistors in the common-emitter configuration, the h^e parameters are the most convenient, as we shall see below. For this reason, and also because they can easily be determined by direct measurement with A.C. (this will likewise be gone into below — see page 25), the h^e parameters are often included in transistor data. It may be added that the various possible sets of parameters are not independent: given one set, one can calculate all the others [7]. h^e parameters for Philips OC 70 and OC 71 transistors at their normal biassing levels are given in *Table I*.

Table I. h^e parameters of Philips OC 70 and OC 71 transistors at normal biassing.

	OC 70 $V_{ce} = -2$ V $I_e = -0.5$ mA	OC 71 $V_{ce} = -2$ V $I_e = -3$ mA
h^e_{11}	2200 ohm	800 ohm
h^e_{12}	9×10^{-4}	5.4×10^{-4}
h^e_{21}	30	47
h^e_{22}	23×10^{-6} mho	80×10^{-6} mho

Characterization of a fourpole by less than four parameters

Various cases arise in which two parameters, or even one, suffice to characterize a fourpole, instead of the four that are generally necessary. This will be made clear by imagining a generator connected to an external resistance R_e, the generator supplying an e.m.f. of E_0 and having an internal resistance of R_0 (see fig. 8). The current delivered is $I = E_0/(R_0 + R_e)$. Since E_0/R_0 is the short-circuited current I_k, we can write:

$$\frac{I}{I_k} = \frac{1}{1 + R_e/R_0}. \quad \ldots\ldots \quad (3)$$

At the terminals of the generator, the voltage $V = IR_e$, so that

$$\frac{V}{E_0} = \frac{R_e/R_0}{1 + R_e/R_0}. \quad \ldots\ldots \quad (4)$$

In *fig. 8*, I/I_k and V/E_0 are plotted as functions of R_e/R_0, a logarithmic scale being employed for both ordinate and abscissa. When plotted in this way, one

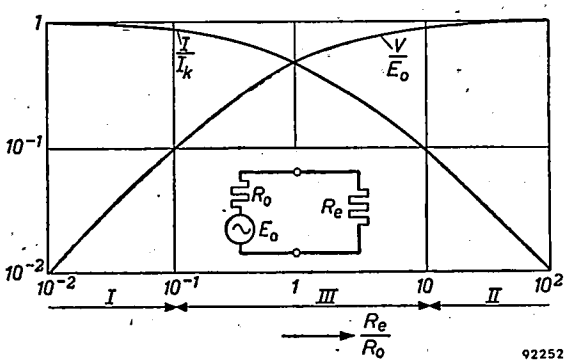


Fig. 8. Current and terminal voltage of a generator (e.m.f. E_0, internal resistance R_0) connected to an external resistance R_e, as functions of the ratio R_e/R_0. The scales of both axes are logarithmic. In region *I* the current approaches the short-circuit value I_k. In region *II* the terminal voltage approaches E_0, the e.m.f. of the generator. In region *III* matters are more complicated.

sees that it is possible to distinguish three regions of R_e/R_0, which are marked in fig. 8 as *I*, *II* and *III*. If R_e is small in comparison with R_0 (region *I*), the *current* is roughly constant and equal to the short-circuit current. If, on the other hand, R_e is large in comparison with R_0 (region *II*), it is the terminal

[6]) Frequently a prime (′) is used to show that the common-emitter configuration is intended, the h parameters without prime being made to relate to the common-base configuration. The drawback of this system is that it prevents the h parameters being used for any sort of fourpole.

[7]) Tables of transformation formulae may be found in: R. F. Shea, Principles of transistor circuits, Wiley & Sons, New York 1953, page 335.

voltage that is roughly constant and nearly equal to the e.m.f. In the middle region (*III*) matters are more complicated.

The four cases where only one parameter need be considered

We shall now apply these simple considerations to a fourpole. At its input the fourpole has a signal source (e.m.f. E_g, internal resistance R_g), and across its output is a resistive load R_l. At the *input side*, therefore, we have a generator (the signal source) connected to an external resistance, this being formed by the *input resistance* r_i of the fourpole with its load R_l (r_i is a function of R_l). At the *output side* we have a generator (the fourpole plus the signal source taken as a whole) connected to an external resistance R_l. The internal resistance of this generator is of course the *output resistance* r_u of the fourpole (r_u is a function of R_g).

It may happen that both the input/source resistance ratio r_i/R_g and the load/output resistance ratio R_l/r_u fall within region *I*. In this case i_1, the input current of the fourpole, is equal to the short-circuit current of the signal source and is thus, as far as the fourpole is concerned, a quantity determined by external conditions. i_2, the output current of the fourpole, is equal to the short-circuit output current, and from (2b) — v_2 now being approximately zero — is equal to $h_{21}i_1$. In this case, therefore, it is only necessary to know h_{21}, the current amplification factor, in order to describe the behaviour of the fourpole [8]). As we shall show by means of an example (pp. 23-24), this case arises when transistors in the common-emitter configuration are directly coupled in cascade. This explains why it is customary in transistor work to place so much emphasis on the current amplification factor; at the same time it is a good reason for including h_{21} among the set of fourpole parameters chosen to characterize the transistor. From the definition of the current amplification factor, viz. $h_{21} = (\partial I_2/\partial I_1)_{V_2}$, it can be seen that in order to arrive at such a set (including h_{21}, that is), I_1 and V_2 must be chosen as independent variables.

The three other cases in which one parameter suffices may be summarized as follows: input in region *II*, output in region *I* (fig. 8); both input and output in *II*; and finally, input in *I* and output in *II*.

Closer investigation of these cases shows up one of the essential differences between transistors, pentodes and triodes in sharp relief. In the case "input in *II*, output in *I*", the input voltage v_1 is a given quantity, being the e.m.f. of the signal source, while the output current is equal to the short-circuit value. Since this latter is given by $i_2 = v_1(\partial I_2/\partial V_1)_{V_2}$, the behaviour of the fourpole is wholly determined by the parameter $(\partial I_2/\partial V_1)_{V_2} = S$, the slope. This is e.g. the case arising with pentodes in the usual grounded-cathode circuits, and it is in fact usual with pentodes to place most emphasis on the slope of the tube. As may be seen from the definition of S, the set of fourpole parameters involving it is obtained by choosing V_1 and V_2 as independent variables.

In the case "both input and output in region *II*", v_1 is equal to the e.m.f. of the signal source, as it is in the previous case, and is again a given quantity. The output voltage has the value appropriate to the open-circuited output, when, in consequence, i_2 is approximately zero; that is, I_2 is constant. Here, therefore, $v_2 = v_1(\partial V_2/\partial V_1)_{I_2}$. The important parameter is here $(\partial V_2/\partial V_1)_{I_2} = \mu$, the voltage amplification factor. The case frequently arises in conventional triode circuits, where, of course, the internal resistance is much lower than with pentodes. It is clear that μ belongs to the set of parameters obtained when V_1 and I_2 are chosen as independent variables.

In the fourth and last straightforward case, that of "input in *I* and output in *II*", $v_2 = i_1(\partial V_2/\partial I_1)_{I_2}$ and hence $(\partial V_2/\partial I_1)_{I_2}$ is the important parameter; it belongs to the set obtained when I_1 and I_2 are chosen as independent variables. This case may arise in a grounded-grid triode circuit.

A case where only two parameters need be considered

Of the cases in which only *two* fourpole parameters are sufficient to describe the behaviour of the fourpole, we shall discuss only that in which the input/source resistance ratio is in region *III* and the load/output resistance ratio in region *I*. It often arises when transistors in the common-emitter configuration are employed in simple amplifiers with resistance-capacity coupling between stages (see page 25). The output is effectively short-circuited ($v_2 \approx 0$) and the output current is therefore $i_2 = h^e_{21}i_1$. The statement that the input/source resistance ratio falls within region *III* expresses the fact that the input resistance of the transistor ($= h^e_{11}$, the value for short-circuited output, since v_2 is approximately zero [9]), see eq. (2a)) is not negligible in comparison with the source resistance R_g. Knowing this, we must calculate the input current from $i_1 = E_g/(R_g + h^e_{11})$. In order therefore to know what we can expect from our transistor, we must know both h^e_{11} and h^e_{21}, these being respectively the input resistance with output short-circuited and the current amplification factor. Again the important parameters are *h* parameters.

Input and output resistance, current gain, voltage gain and power gain

Calculation and graphical representation

It will have become clear from the foregoing that r_i and r_u, the input and output resistances of a fourpole, have an important bearing on its behaviour.

[8]) Strictly speaking, this reasoning need not always be valid: its validity can be upset by the feedback effect expressed by h_{12}. We do not want to interrupt the argument at this point for the sake of this refinement, but shall return to it later on.

[9]) The reservation made in footnote [8]) must again be made here.

If the fourpole is specified by its h parameters (or by any other set of parameters), it is an easy matter to work out r_i and r_u. r_i is the equivalent resistance of the twopole shown in *fig. 9a*. From the figure we see that

$$v_2 = -i_2 R_l . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

By eliminating i_2 and v_2 from this equation with the help of (2a) and (2b) we obtain r_i $(= v_1/i_1)$ as a function of the load resistance R_l [10]). The resulting formula is given in *Table II*, and a graphical representation for a practical case (a Philips OC 71 transistor in the common-emitter configuration) is given in *fig. 10a*.

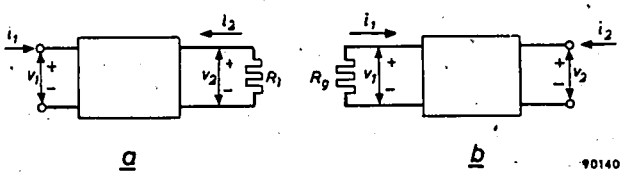


Fig. 9. *a*) Circuit for calculating the input resistance r_i of a fourpole having a resistance R_l as load. *b*) Circuit for calculating the output resistance r_u of the same fourpole when the generator has an internal resistance of R_g.

(2b) gives r_u $(= v_2/i_2)$ as a function of the source resistance R_g (see Table II and fig. 10*b*).

Since we have in mind the employment of the fourpole as an amplifier, we have a particular interest in the current, voltage and power gains, denoted by A_i, A_v and G respectively. A_i $(= i_2/i_1)$ is obtained by eliminating v_2 from (2b) and (5). A_v $(= v_2/v_1)$ is obtained by eliminating i_1 and i_2 from (2a), (2b) and (5). Conventions as to the sign of voltages and currents being duly observed, the power absorbed by the fourpole at the input side is $P_1 = i_1 v_1$, and the power delivered at the output side $P_2 = -i_2 v_2$. Thus the power gain is $G = -A_i A_v$. The formulae for A_i, A_v and G are given in Table II, and their graphical representations for the OC 71 in the common-emitter configuration appear in fig. 10*a* as functions of R_l.

The relative complexity of the formulae in Table II is chiefly due to the feedback from the output to the input, which is expressed by h_{12}. Formulae for the case $h_{12} = 0$, which also appear in the table, show the simplification that would result were this effect absent. In the second set of formulae input and output resistances are constants of the transistor, i.e. r_i is no longer dependent on the load, nor r_u on the signal source resistance.

Table II. Left column: some important fourpole quantities expressed in terms of the h parameters and the load and source resistances R_l and R_g. p is an abbreviation for $h_{12}h_{21}/h_{11}h_{22}$. The quantities thus given are, in order: input and output resistance, r_i and r_u; current, voltage and power gain, A_i, A_v and G; and finally G_{max}, the maximum value of the power gain. Right column: formulae for the same quantities in the case where there is no feedback from output to input (i.e. when h_{12}, the reverse voltage amplification factor, is zero).

$r_i = h_{11}\left(1 - p\dfrac{h_{22}R_l}{1+h_{22}R_l}\right)$	$r_i = h_{11}$
$r_u = \dfrac{1}{h_{22}\left(1 - p\dfrac{1}{1+R_g/h_{11}}\right)}$	$r_u = \dfrac{1}{h_{22}}$
$A_i = \dfrac{h_{21}}{1+h_{22}R_l}$	$A_i = \dfrac{h_{21}}{1+h_{22}R_l}$
$A_v = \dfrac{1}{h_{12}\left\{1 - \dfrac{1}{p}\left(1+\dfrac{1}{h_{22}R_l}\right)\right\}}$	$A_v = -\dfrac{h_{21}R_l}{h_{11}(1+h_{22}R_l)}$
$G = \dfrac{h_{21}}{h_{12}} \cdot \dfrac{ph_{22}R_l}{(1+h_{22}R_l)\{(1-p)h_{22}R_l + 1\}}$	$G = \dfrac{h_{21}^2}{h_{11}h_{22}} \cdot \dfrac{h_{22}R_l}{(1+h_{22}R_l)^2}$
$G_{max} = \dfrac{h_{21}^2}{h_{11}h_{22}} \cdot \dfrac{1}{(1+\sqrt{1-p})^2}$	$G_{max} = \dfrac{h_{21}^2}{4h_{11}h_{22}}$
(G_{max} occurs when $h_{22}R_l = 1/\sqrt{1-p}$)	(G_{max} occurs when $h_{22}R_l = 1$)

The output resistance r_u is the equivalent resistance of the twopole shown in fig. 9*b*. Here we may write:

$$v_1 = -i_1 R_g . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Elimination of i_1 and v_1 from equations (6), (2a) and

Inferences to be drawn from the graphs

Let us imagine a directly-coupled cascade of three transistors, T_1, T_2 and T_3, of type OC 71, in the common-emitter configuration (*fig. 11a*), biassed to an operating point of $I_c = -3$ mA and

[10]) We shall confine ourselves to the cases where R_l and R_g are real (i.e. pure resistances). The h-parameters are also real at the low frequencies here assumed, and for this reason we shall also find that r_i and r_u have real values.

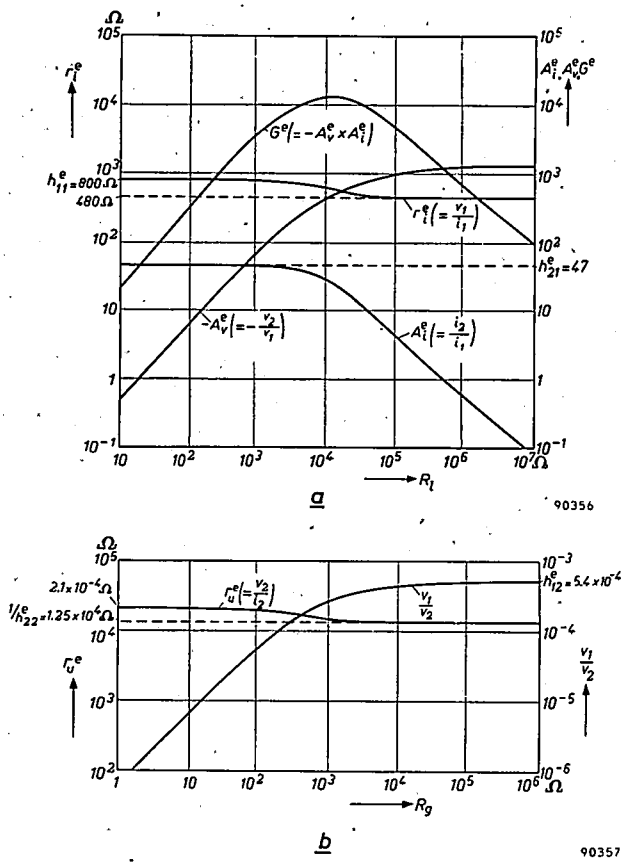


Fig. 10. Some important quantities plotted for the Philips OC 71 transistor in the common-emitter configuration (hence the e indices to the upper-right of symbols). All axes are on logarithmic scales. D.C. biassing (operating point): $I_c = -3$ mA, $V_{ce} = -2$ V.
a) Input resistance r_i^e, current gain A_i^e, voltage gain A_v^e ($-A_v^e$ is plotted here because, on account of the accepted convention, A_v^e is negative) and power gain G^e, as functions of the load resistance R_l.
b) Output resistance r_u^e and the reverse voltage amplification factor v_1/v_2, as functions of the source resistance R_g.

$V_{ce} = -2$ V. T_2 is therefore "sandwiched" between two other transistors. $r_u^e(T_1)$, the output resistance of T_1, is the source resistance $R_g(T_2)$ for T_2. Since $r_u^e(T_1)$ lies somewhere between $2.1 \times 10^4\ \Omega$ and $1.25 \times 10^4\ \Omega$ (see fig. 11*b*) — its exact value depends on the source resistance $R_g(T_1)$ — it follows from fig. 11*d* (hatched area) that $r_u^e(T_2) = 1.25 \times 10^4\ \Omega$. And since $r_u^e(T_2) = R_g(T_3)$, we find from fig. 11*f* that $r_u^e(T_3)$ has the same value of $1.25 \times 10^4\ \Omega$.

$r_i^e(T_3)$, the input resistance of T_3, lies somewhere between 800 Ω and 480 Ω (see fig. 11*g*), its exact value depending on its load, $R_l(T_3)$. Fig. 11*e* (hatched area) shows that, since $r_i^e(T_3) = R_l(T_2)$, then $r_i^e(T_2)$ is 800 Ω. In the same way, making use of $r_i^e(T_2) = R_l(T_1)$, we find from fig. 11*c* a value of 800 Ω for $r_i^e(T_1)$.

We learn from fig. 11*c* that the input resistance and current gain of T_1 have the values appropriate to short-circuiting of the output, in other words, they are equal to h_{11}^e and h_{21}^e respectively. Consequently these two h^e parameters are all we need to know about T_1. If data on the signal source are available, we shall be in a position to calculate the input current of T_1; this, multiplied by h_{21}^e (the current amplification factor), will give us the output current. The output current of T_1 is the input current of T_2. As we learn from fig. 11*e*, the current amplification of the second transistor T_2 is also h_{21}^e (value of $A_i^e(T_2)$ where curve is horizontal, see also fig. 10*a*); multiplying by this quite straightforwardly (here it is not necessary to know h_{11}^e) gives us the output current of T_2, which is, in turn, the

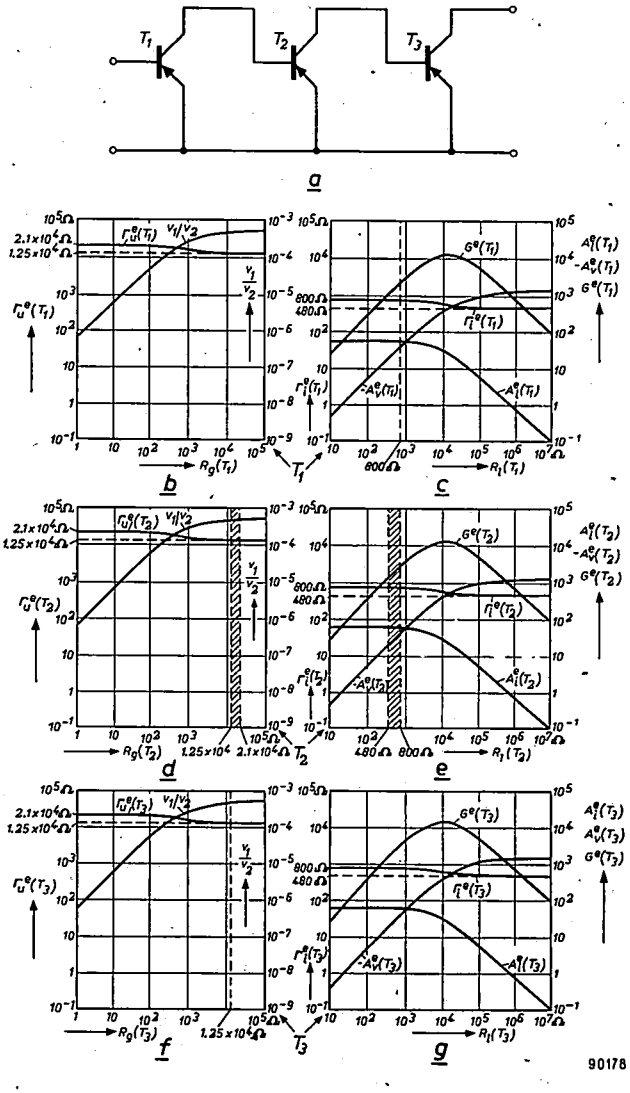


Fig. 11. Illustration of the situation obtaining when, as in (*a*), three transistors of type OC 71, T_1, T_2 and T_3, each in the common-emitter configuration, are directly coupled in cascade. The curves in (*b*), (*d*) and (*f*) are identical with those in fig. 10*b*, and relate to the outputs of T_1, T_2 and T_3 respectively. The curves in (*c*), (*e*) and (*g*) are identical with those in fig. 10*a*, and relate to the three inputs. The six graphs are annotated with such statements about the situation at the input or output in question as can be made without knowledge of the source resistance of T_1 or the load resistance of T_3.

input current of T_3. If, in addition, we know that $R_l(T_3)$ is lower than 1000 Ω, say, as in many cases it will be, fig. 11*g* shows that the current amplification will once again be equal to h_{21}^e. Multiplying by this once more, we obtain the output current of T_3. If, on the other hand, $R_l(T_3)$ is so large that A_i^e falls much below h_{21}^e, we should have to know h_{22}^e as well as h_{21}^e in order to calculate first A_i^e (using the formula given in Table II) and then the output current of T_3.

When resistance-capacity coupled amplifying stages are used (*fig. 12*), coupling resistors R_k occur across the outputs of the successive transistors. For simplicity we shall assume the capacitors to have so large a value that they offer no obstacle to alternating current. This means that the load resistance formed by the transistor of the following stage lies in parallel with the coupling resistor. The resultant load resistance is therefore diminished, and we can now say with still more certainty that the current gain and input resistance of the stage have the values appropriate to a short-circuited output, namely h_{21}^e and h_{11}^e. This being so, we can work out the output current of T_1 in the fashion explained above; but in the present case that current will not have the same value as the input current of T_2. In order to obtain the input current of T_2, we must calculate how the output current of T_1 is divided between the coupling resistor and the input resistance of T_2; for this it is necessary to know the latter, i.e. to know h_{11}^e. The same, of course, applies to the calculation of the input current of T_3, always assuming that $R_l(T_3)$ is less than 1000 Ω.

We could have reached the same conclusions by making use of fig. 8 which, with the aid of the numerical values found above, would allow us to investigate the input/source resistance ratio and the load/output resistance ratio of the three transistors.

The two examples just dealt with demonstrate the importance of both the current amplification factor, and the input resistance with the output short-circuited.

There is a further important point that can be clarified by the graphs of fig. 10: it is possible to measure the h^e parameters directly and without difficulty at low frequency, say 1000 c/s (at which frequency the static characteristics are still obeyed). In principle some of the measurements should be carried out in a circuit with open input and others with short-circuited output. An actual break in the input loop, like actual short-circuiting of the output, would, of course, make it impossible to apply the necessary D.C. biassing. However, it is necessary neither to make R_g particularly large nor R_l particularly small in order to have the effect of an open input and a short-circuited output, respectively. It will be seen from fig. 10 that for an OC 71 transistor it would suffice for R_g to be greater than 5000 Ω and for R_l to be less than 1000 Ω. Of course, when the parameters of an unknown transistor are being measured, these limits are also unknown in the first place. In such circumstances the measurements are repeated for increasing values of R_g and decreasing values of R_l, up to the point where v_1/v_2 and i_2/v_2 on the one hand and v_1/i_1 and i_2/i_1 on the other cease to change in value.

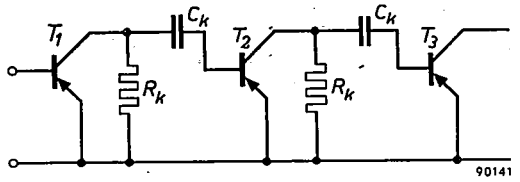


Fig. 12. Three transistors coupled by the resistors R_k and the capacitors C_k.

Internal feedback

In footnote [8]) a reservation was made concerning the conclusion that $i_2 = h_{21}i_1$ in the case where both the input/source resistance ratio and the load/output resistance ratio fell within region *I* (see fig. 8); in accordance with this conclusion — it was stated in the text — the current amplification factor was sufficient to describe the fourpole. The difficulty referred to in the footnote is the following. Suppose that R_g and R_l have values such that the input/source and the load/output resistance ratios do in fact both fall within region *I*. Now, if R_l (= R_c in fig. 8 when we consider the situation at the output) is reduced, the value of the load/output resistance ratio R_l/r_u (= R_c/R_0) is shifted further into region *I*, and consequently the value of i_2 is unaffected by the change. But r_i is a function of R_l. This is a consequence of the internal feedback, for if feedback is absent, i.e. if $h_{12} = 0$, it follows from the formula for r_i in Table II that $r_i = h_{11}$ = constant. If now r_i (= R_c in fig. 8 when we consider the situation at the input) increases as R_l decreases, then the value of the input/source resistance ratio r_i/R_g (= R_c/R_0) is shifted to the *right* in fig. 8. It is therefore conceivable that, as R_l approaches zero, the input/source resistance ratio moves out of region *I*. If so, i_1 ceases to be the short-circuit current of the signal source; it takes on a lower value. The value we would obtain for i_2 by multiplying the short-circuit current of the signal source by the current amplification factor h_{21} would therefore be too high. For transistors in the common-emitter configuration, r_i does indeed increase as R_l decreases (see the example of the OC 71 in fig. 10*a*), and one might justifiably think that the conclusion $i_2 = h_{21}i_1$ (for input/source and load/output resistance ratios both within region *I*) would be erroneous at least in some cases. If one goes deeper into the matter, however, it turns out that, at least for junction transistors, the conclusion is always valid for the case of both input/source and load/output resistance ratios lying within region *I*.

Similar considerations apply to the other cases dealt with on page 22. Strictly speaking, classification with the aid of fig. 8 of the various situations arising in fourpoles is only really possible if the feedback effect is neglected. With a vacuum tube, in the usual type of circuit (viz. common-cathode circuit), there is no feedback effect at low frequencies. This and

the fact that the input current in such a circuit is zero are the reasons why the treatment of tubes as linear, active, network elements is very much simpler than that of transistors.

Power gain and figures of merit of amplifiers

In general, an amplifier is employed whenever it is desired to have a power greater than is available without an amplifier. In most cases, however, the power gain G is by no means a good criterion for evaluating the performance of an amplifier in given conditions. The inadequacy of G becomes particularly clear in the extreme case of tube amplifiers (here we have in mind receiving tubes in the conventional common-cathode circuit). The input current is then negligibly small, and so in consequence is P_1, the power absorbed at the input. Since P_2, the power delivered, is finite, G is almost infinite. This certainly does not mean that any tube in any circuit is necessarily a good amplifier: whether the power delivered by the tube is greater than that which the signal source would have delivered without the intermediacy of the tube, remains an open question. And, in fact, reference is hardly ever made to power gain in connection with tubes. With transistors it is a different matter; since their input current is never zero, it is reasonable enough to talk about the power gain they give. Accordingly, in fig. 10*a*, G has been plotted as a function of R_l, this being a curve that may sometimes be useful in designing amplifiers. However, the G of transistors is still not a sure guide to the merits of the amplifier, for the question posed above remains unanswered. P_2 is given by GP_1; should P_1 be low because the transistor is badly matched to the signal source, it is quite possible that a higher value of P_2 might be made available by having better matching at the input, even though G should be lower.

A better figure of merit is arrived at by comparing the behaviour of an amplifier inserted between a signal source S and a load R_l, with that of an ideal matching transformer; in other words, by dividing the power that the amplifier delivers to R_l by the power that could be fed to R_l under the most favourable conditions, but without an amplifier. The figure thus obtained is called the "operating gain" or, alternatively, "transducer gain". We would invite special attention to the fact that *two* circuits are involved in the definition of operating gain, one circuit with an amplifier and one without. In fact, therefore, it is not a question here of a "gain of power" in the usual sense of the words — in the sense in which they are used in the term "power gain", for example. The sense is rather that of "advantage".

The concept of operating gain can reasonably be applied to any kind of amplifying device, including tubes, provided the amplifying device is considered in conjunction with the signal source and the load; it is not a quantity that can be stated for an amplifying device standing by itself. Indeed, the term itself says as much.

For a given signal source, the operating gain is at a maximum when amplifier and load are matched to each other. In the American literature, operating gain with matched output is called "available gain". It is possible to go a step further and also match the input: we then get the "maximum available gain". It is not difficult to see that maximum available gain is equal to G_{max}, the maximum power gain (see Table II), and is therefore a quantity that can be stated for an amplifying device standing alone, including a single transistor. For the OC 71 transistor in the common-emitter configuration it is approx. 10 000, i.e. approx. 40 dB.

To obtain the maximum available gain from a transistor is not impossible in practice, but doing so usually implies the use of matching transformers at input and output. The maximum available gain provided by tubes is almost infinite, but in view of the impossibility of matching the input, the fact is of no practical importance. In transistor circuits transformers will frequently be ruled out by considerations of space and price, the employment of one or more extra transistors being preferred.

Clearly, the comparison circuit need not necessarily be that in which the amplifying device is replaced by an (ideal) transformer. Sometimes the signal source directly connected to the load is chosen for this purpose. The figure of merit thus obtained is termed the "insertion gain". Here, as in "operating gain", the word "gain" has a sense different from the usual one.

The properties of a transistor as a linear active network element change with its biassing. This matter will be further investigated in a subsequent article, with the aid of an equivalent circuit whose elements are directly related to the physical functioning of the transistor (a "physical equivalent circuit"). In the same article the opportunity will be taken of discussing transistor equivalent circuits more generally. The temperature effects so important in transistor work will be dealt with in connection with the biassing in a third and final article.

Summary. The properties of a transistor at low frequencies can be expressed by pairs of families of curves, many different choices of these being possible. A transistor, like other network elements with three terminals (e.g. vacuum tubes), can be turn-

ed into a fourpole in three different ways, these being known as the common-emitter, common-base and common-collector configurations. Parallels with tube circuits are discussed. Where small signals only are involved, an active fourpole can be regarded as a linear active network element that can be characterized by four fourpole parameters. In simple cases the so-called h parameters are frequently selected for characterizing transistor behaviour. These parameters express the input voltage and output current in terms of input current and output voltage. The usefulness of h parameters is illustrated by examples, use being made of graphs in which input resistance and current gain are plotted as functions of load resistance, and output resistance is plotted as a function of source resistance. For transistors, unlike vacuum tubes in normal circuits, the concept of power gain is a usable one, but it is inadequate in itself as a criterion for the performance of an amplifier. A better guide is provided by "operating gain".

Two subsequent articles will deal with the effect of the biassing on the properties of transistors, and with temperature effects in connection with the biassing.

BOTTLING OF INJECTION FLUIDS



The Philips-Roxane Company markets various kinds of injection fluids [1]). The filling of the phials takes place under conditions of strict sterility, being carried out in a cupboard in which bacteria-free air is kept at a pressure slightly above atmospheric. Bacteria that may still find their way inside are killed by radiation from ultra-violet lamps fitted in the top of the cupboard. The phials are filled from a tube, which can be seen in the background. In the foreground the phials are being capped.

[1]) One of them being influenza-virus vaccine; see A. J. Klein and E. Hertzberger, Philips tech. Rev. **12**, 273-282, 1950/51.

PRE-MAGNETIZATION OF THE CORE OF A PULSE TRANSFORMER BY MEANS OF FERROXDURE

by H. G. BRUIJNING and A. RADEMAKERS.

621.314.2.073.3:621.318.124:621.374.3

Among the outstanding results yielded by recent solid-state research are two groups of non-metallic magnetic materials, ferroxcube and ferroxdure, which have both been the subject of several articles in this Review. Ferroxcube is a soft magnetic material with low losses even at high frequencies. Ferroxdure is a hard magnetic material, very suitable for permanent magnets. A combination of the two materials is found in new pulse transformers, developed for magnetron modulators. These transformers have a ferroxcube core, which is pre-magnetized by a ferroxdure magnet.

In electrical engineering and electronics we often find transformers whose primary or secondary current is entirely composed of pulses in the same direction. Some examples are: transformers feeding half-wave rectifiers, ignition coils for internal-combustion engines and the pulse transformers feeding radar transmitters. The core of such transformers is unfavourably loaded: the magnetic induction, instead of oscillating between $-B_{max}$ and $+B_{max}$, varies only between 0 and B_{max}, so that one half of the magnetization curve is not used at all. There is a means, however, to employ both halves, viz. by pre-magnetizing the core. We shall demonstrate in the following that this can be successfully effected by means of a *ferroxdure* permanent magnet [1]).

In this article we shall confine ourselves to the application of this principle in pulse transformers for radar transmitters.

Working principle of a radar transmitter

A radar transmitter radiates periodically wave trains of large power and short duration. The peak power may be of the order of hundreds of kW, the duration of each wave train 1 μsec, and the repetition frequency 1000 c/s. In order to operate the oscillator tube (as a rule a magnetron) intermittently, it is fed with an anode voltage of approximately square waveform.

Fig. 1 shows a circuit of the type commonly employed to generate such anode-voltage pulses. The thyratron *Th* being not ignited, the direct-voltage source *E* supplies a charging current flowing through the choke coil *L*, the pulse-shaping network *P* and the pulse transformer *T*. The pulse-shaping network *P*, which may be a lumped transmission line composed of coils and capacitors (delay line) is thereby charged. The magnetron *M* is connected to the secondary winding of *T*. By igniting the thyratron, the points *A* and *C* are suddenly short-circuited (short-circuiting of the D.C.-source is prevented by the choke *L*). This produces a square

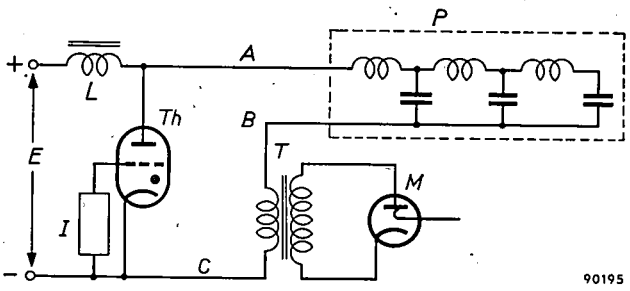


Fig. 1. Circuit for supplying the anode-voltage pulses for a radar transmitter. *E* direct-voltage source. *L* choke coil. *Th* thyratron. *I* ignition device for the thyratron. *P* pulse-shaping network (delay line). *T* pulse transformer. *M* magnetron.

voltage pulse across the primary of pulse transformer *T*, provided that the characteristic impedance of the delay line *P* is equal to the impedance between the points *B* and *C*. The duration of this pulse is equal to the time required by a wave front to travel through the network *P* and back again. During this pulse the energy stored up in *P* is transferred, via the pulse transformer, to the magnetron and partly radiated as high-frequency energy. At the end of each pulse the thyratron current falls to zero and the sequence can be repeated.

The pulse transformer

Frequency response

Magnetrons handling large powers require a high voltage (several ten-thousands of volts). This is produced by the pulse transformer *T*, which steps up a lower voltage supplied to the primary. A typical example: secondary 28 kV, 35 A; primary 7 kV, 140 A (peak values).

[1]) J. J. Went, G. W. Rathenau, E. W. Gorter and G. W. van Oosterhout, Philips tech. Rev. **13**, 194-208, 1951/52.

Fig. 2a shows the primary voltage v_1 as a function of the time t, and fig. 2*b* the desired form of the magnetic induction B in the transformer core during

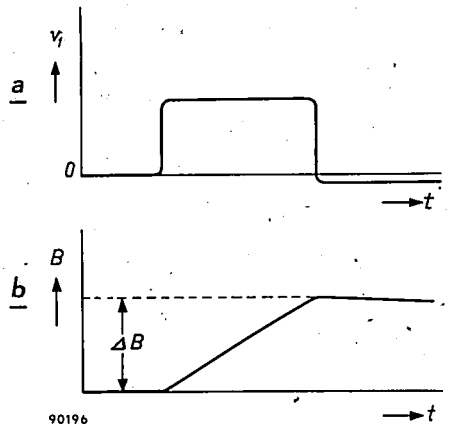


Fig. 2. *a*) Primary voltage v_1 of the pulse transformer T in fig. 1, versus time t. *b*) Magnetic induction B in the transformer core versus time.

the same period. The transformer must therefore transfer as faithfully as possible the entire frequency spectrum of v_1; this spectrum is shown (for relatively far too wide pulses, *fig. 3a*) in fig. 3*b*. The frequency-response curve of the transformer, however, will never be completely flat. Owing to the finite self-inductance of the primary, the induction B will not have the form of fig. 2*b*, but there will be a loss at the low frequency side, whilst the highest frequency that is adequately transformed will be determined by the required length of wire per turn. The frequency-response curve has, for example, a shape like that shown in fig. 3*c*, with the result that the secondary frequency spectrum will be of the form shown in fig. 3*d*. In the shape of the secondary voltage the loss of low-frequency components manifests itself as a "sagging" of the waveform (see fig. 3*e* and the oscillograms *fig. 4*) so that, in the wave train transmitted by the magnetron during a single pulse, the transmitted power diminishes accordingly. The loss of high-frequency components reduces the steepness of the leading and trailing edges of the pulse, so that the transmitted wave trains are less sharply defined. The first effect is more apparent if, for a given transformer core, too few windings have been used (self-inductance too low), and the second effect predominates if the total wire length and hence the number of turns is too large. A compromise is therefore to be sought.

Use of a ferroxcube core

The compromise attainable can be favourably influenced by suitable choice of the core material. A material of high permeability is advantageous in that it requires relatively fewer turns to give a specified self-inductance. A high saturation induction is favourable in view of the small diameter required for the core (small length per turn and hence shorter total wire length). If at the same time the eddy-current and hysteresis losses are low, the transformer will have a high efficiency.

High permeability, low losses and a not too low saturation induction are favourably combined in certain types of ferroxcube. From an engineering point of view, ferroxcube has the advantage over a metal core in that the former can be very easily composed from a few solid blocks; for a metal core only thin-rolled nickel-iron strip, wound into a ring, is suitable for this purpose. Ferroxcube moreover makes it possible to give different parts of the magnetic circuit different diameters, which is found to be of great advantage with pulse transformers; a ring

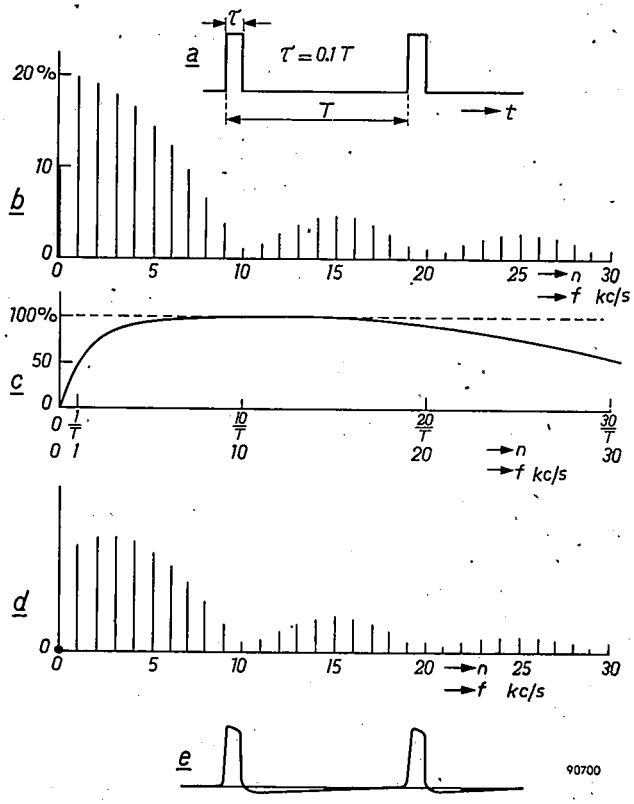
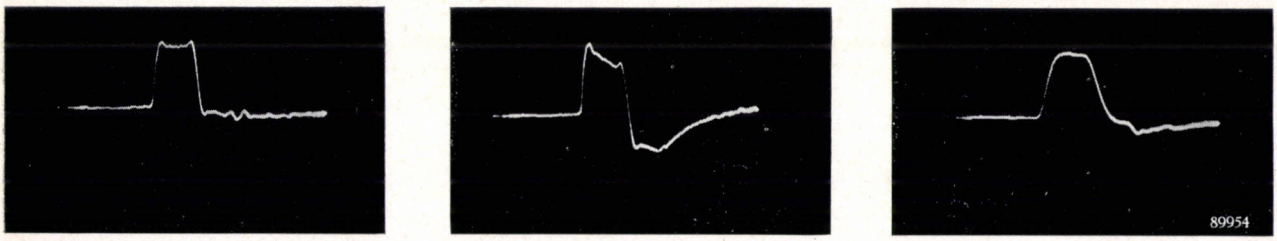


Fig. 3. *a*) Rectangular pulses of duration τ and repetition frequency $1/T$. With a view to the clarity of figs. (*b*) and (*d*), the pulses are shown relatively far too wide ($\tau/T = 0.1$ instead of approx. 0.001).
b) Spectrum of the pulse train (*a*). The amplitudes of the harmonics have been plotted vertically as a percentage of the pulse height; horizontally plotted are the order n and the frequency f of the harmonics, for the case $\tau = 10^{-4}$ sec, $T = 10^{-3}$ sec.
c) Frequency response of a pulse transformer.
d) Spectrum of the secondary voltage pulses.
e) The secondary pulses are distorted owing to the shape of curve (*c*).
(For the case of $\tau = 10^{-6}$ sec, $T = 10^{-3}$ sec, which is closer to normal radar usage, the scales of n and f should be multiplied by 100, and the line density in (*b*) and (*d*) should be 100× greater.)



a *b* *c*

Fig. 4. Oscillograms of the secondary voltage of a pulse transformer whose primary voltage consists of a series of nearly rectangular pulses (duration 1 μsec) and whose secondary winding is resistance-loaded.
a) Suitable shape of secondary pulses.
b) The too small primary self-induction causes a "sagging" pulse (loss of low-frequency harmonics).
c) Too long a wire length reduces the steepness of leading and trailing edges (loss of high-frequency harmonics).

wound from metal strip must necessarily retain a uniform diameter.

Ferroxcube 3C2, whose magnetization curve is shown in *fig.* 5, is particularly suited to this purpose. The saturation induction is about 0.5 Wb/m² (= 5000 gauss), corresponding to a magnetic field strength $H \approx 16$ kA/m (~ 200 oersted). If the material is to be employed at optimum efficiency, the variation ΔB should be enabled to occupy a long, steep section of the *B-H* curve. Suppose that ΔB be 0.3 Wb/m² and that this variation take place between the points *O* and *P* (fig. 5), then the relative permeability μ_r will be about 300. If the number of turns is so reduced that $\Delta B = 0.5$ Wb/m², and the curve *OQ* is chosen, then the increased yield of high frequencies (because of the smaller wire length) is offset by a loss of low frequencies, because μ_r drops to about 25, the self-inductance thus becoming considerably smaller.

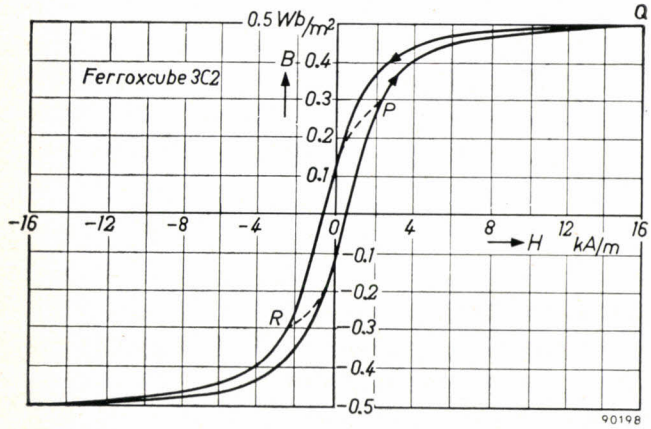


Fig. 5. *B-H* curve of ferroxcube 3C2. The curve *OP* corresponds to a large μ_r (= 300), but ΔB is only 0.3 Wb/m². If the curve *OQ* is chosen, then $\Delta B = 0.5$ Wb/m², but μ_r is reduced to 25. The condition with a pre-magnetization corresponding to −0.3 Wb/m² (point *R*) is a great deal more favourable: for the curve *RP*, $\Delta B = 0.6$ Wb/m² and $\mu_r = 300$.

Pre-magnetization

Conditions can be greatly improved if the steep part of the *B-H* curve below the *H*-axis is likewise employed. If, by applying a pre-magnetization of −0.3 Wb/m², *R* (fig. 5) is taken as the starting point instead of *O*, then the curve *RP*, corresponding to $\Delta B = 0.6$ Wb/m² and $\mu_r = 300$, is available. The application of pre-magnetization is therefore very profitable.

The question remains: in what way can pre-magnetization be most effectively established? The transformer core might be given a third winding, through which a constant direct current could be passed of such intensity that the core would be adequately pre-magnetized. This method has the advantage that it is possible to retain a *closed* magnetic circuit, so that the large permeability of the core material can be fully utilized. It has the drawback, however, of requiring a separate source of D.C.; moreover, the excitation circuit must have such a high impedance that it absorbs no appreciable amount of pulse energy.

A better solution is pre-magnetization by means of a *permanent magnet*, either of magnet steel or of ferroxdure. The latter has the advantage that it can pass an alternating flux without causing any appreciable eddy-current losses, owing to the high resistivity of this material.

Against the advantage that a permanent magnet does not require a D.C. source there is the drawback that the magnetic circuit must be interrupted for incorporating the magnet, thus rendering it more difficult to obtain a given self-induction. Particularly in the cause of ferroxdure, with its small reversible permeability, some difficulties in this respect would be expected. This drawback, however, is found to be a minor one: thanks to the high coercive force of

ferroxdure, a thin slab of this material is sufficient, so that the air-gap to be made in the transformer core can be narrow and consequently the decrease in self-induction remains moderate. We shall return to this later in this article. Let us first examine how the ferroxdure reacts to a current pulse through the transformer.

Behaviour of ferroxdure subjected to a pulse-shaped demagnetizing field

First of all we shall consider the general case of a pulse transformer with a core, pre-magnetized by a steel permanent magnet. The *B-H* curve of magnet steel is represented in *fig. 6*. The initial condition of the magnet is characterized by a point on the part of the *B-H* curve in the second quadrant, e.g. point *1*.

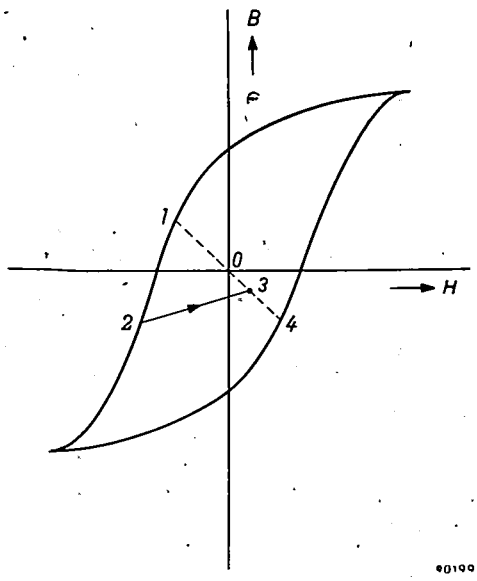


Fig. 6. *B-H* curve of a magnet steel. *1-0-4* working line of a magnet made of this steel, pre-magnetizing a pulse transformer. *1* initial working point. During a pulse the flux is reversed and point *2* is reached. At the end of the pulse, point *3* becomes the new working point. The pre-magnetization has thus taken on the wrong polarity.

If now the transformer is operated by a pulse of such amplitude and direction that it changes the sign of the induction, then the condition of the magnet will be temporarily represented by the point *2*, which lies equally far below the *H*-axis as *1* lies above it. The aim is obviously to have the initial condition restored after the pulse, but it will be readily appreciated that this is by no means certain. As soon as the pulse disappears, the working point traverses the line from *2* to *3*; the terminal point *3* being situated on the "working line" *1-0-4* of the system [2]. The line *2-3* has only a slight upwards slant (the "reversible permeability" of the magnetic material), so that *3* may be very well situated below the *H*-axis. This would mean that the pre-magnetization had not only changed its value, but also taken on the wrong sign, which, of course, has to be prevented at all costs.

With ferroxdure, however, matters are quite different. The reversible permeability of ferroxdure I is only little less than the slope of the *B-H* curve, so that the return curve lies closely along and nearly parallel with the *B-H* curve. The induction of the material, moreover, can be given a considerable negative value without the risk of losing the magnetization on the return path. An idealized representation of this behaviour, which we shall initially use in the reasoning which follows, is shown in *fig. 7*.

Approximated by straight lines, the magnetization J of ferroxdure I is plotted against H in fig. 7*a*. With increasing intensity of the demagnetizing field, J remains constant as long as H does not exceed the value ${}_J H_c$, but when H does exceed this value, J is entirely reversed. Fig. 7*b* shows the corresponding *B-H* curve: $B = J + \mu_0 H$, in which μ_0 is the permeability of free space ($= 4\pi \times 10^{-7}$ Wb/Am). When $\mu_0 H$ is plotted as abscissa on the same scale as B, as is done in fig. 7*b*, then the lines

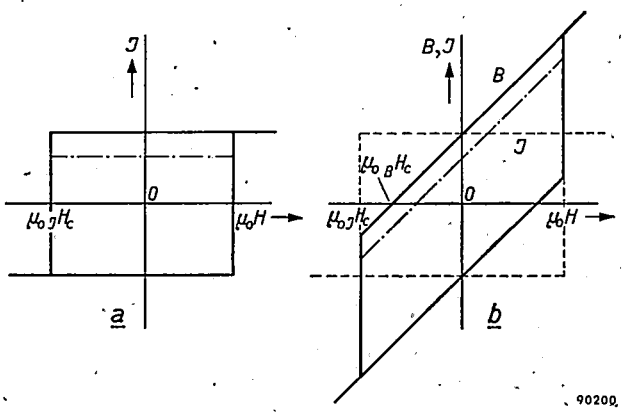


Fig. 7. *a*) Magnetization J, and (*b*) induction B of ferroxdure, as functions of $\mu_0 H$ (idealized). The value of J is shown as the dotted line in (*b*). The chain-dotted lines in both figures apply to a less strongly magnetized ferroxdure.

run at an angle of 45°. Only the upper slanting line is of interest to us here. It intersects the $\mu_0 H$-axis at a point $\mu_0 {}_B H_c$ (which, in ferroxdure, has a value significantly different from $\mu_0 {}_J H_c$). The working point can be shifted up and down along the upper sloping line of fig. 7*b* in a completely reversible manner. A condition of weaker magnetization is represented in both diagrams by a chain-dotted line, which may likewise be reversibly followed, and for which the same limiting value ${}_J H_c$ of the demagnetizing field applies.

[2]) Further in this article the working line will be more fully explained. See also A. Th. van Urk, Philips tech. Rev. 5, 29-35, 1940.

It should be observed that in this idealized picture the relative permeability $\mu_r = 1$, whereas the actual value for ferroxdure lies between 1.1 and 1.4. We shall deal with the actual B-H curve later in this article.

In our further study of the subject we shall assume that the designer of the pulse transformer has chosen a given primary self-induction L_1 and a given number of primary turns n_1; this also fixes the magnetic resistance (reluctance) R of the magnetic circuit, since if I_m is the magnetizing current and Φ_m the resulting alternating flux, the relations $I_m n_1 = \Phi_m R$ and $L_1 I_m = n_1 \Phi_m$ are valid, so that

$$R = \frac{n_1^2}{L_1} \quad \ldots\ldots\ldots \quad (1)$$

R is the sum of the magnetic resistance R_{fxc} of the ferroxcube part of the circuit and of the magnetic resistance R_s of the air gap to be incorporated in the circuit to accomodate the ferroxcube magnet:

$$R = R_{fxc} + R_s .$$

For R_s we may put:

$$R_s = \frac{d}{\mu_r \mu_0 A_s},$$

d being the width of the gap (equal to the thickness of the ferroxdure slab) and A_s the area of the ferroxcube surfaces bordering the gap (A_s will as a rule be a good deal larger than the core diameter A_k of the ferroxcube within the coil). μ_r approximates to unity, both for ferroxdure and for air. Therefore, irrespective of the ferroxdure filling of the gap, we may put:

$$R_s = \frac{d}{\mu_0 A_s}. \quad \ldots\ldots\ldots \quad (2)$$

We shall assume here that not only R is fixed ($= n_1^2/L_1$, see (1)), but also the distribution of R into R_{fxc} and R_s, and thus also R_s itself. It follows from (2) that then also the ratio d/A_s ($= \mu_0 R_s$) is fixed.

We shall now attempt to determine the smallest admissible values of d and A_s, the smallest, that is, which involve no risk of permanent demagnetization. That this risk decreases with greater values of d and A_s (maintaining the same ratio), is attributable to two causes:

1) The larger the ferroxdure volume, the weaker the necessary pre-magnetization, thus allowing of a larger permissible negative B, according to fig. 7b.
2) The larger the area A_s, the smaller the variation of B resulting from a given flux variation.

These two factors fairly well define the critical condition where the risk of permanent demagnetization has to be seriously considered. It may be added here that, in practice, operation near this critical limit is not advisable. In case of certain faults the back flux may assume values considerably above normal, so that a certain safety margin is always necessary.

The faults in question are those whereby the magnetron fails to oscillate. The pulse transformer will than operate virtually without load and the earlier mentioned condition that the delay line (P, fig. 1) should be in series with a resistance equalling its own characteristic impedance is no longer fulfilled. As a result of this, the magnetizing current becomes many times (e.g. 40×) greater than normal. Fortunately, owing to the resulting saturation of the ferroxcube, the induction B_k in the core increases at a far lower rate than the current; however, the induction rises to the extent that a safety factor of 2-3 is indispensable to avoid demagnetization of the ferroxdure. This safety factor is defined as $(\Delta B_k)_{max}/2B_0$, $(\Delta B_k)_{max}$ being the maximum change in B_k that does not yet cause any demagnetization.

Let us first consider the case of an air-gap completely filled with ferroxdure (*fig. 8a*) and then that of a partly filled gap (fig. 8b).

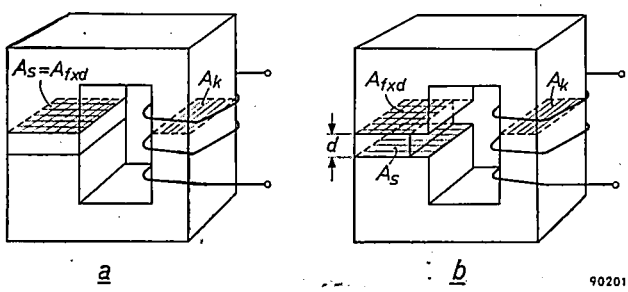


Fig. 8. Simplified shape of the ferroxcube core of a pulse transformer whose gap is filled completely (*a*), and only partly (*b*) with a pre-magnetizing permanent magnet of ferroxdure.

Completely filled gap

Since the pre-magnetizing flux Φ_0 may be assumed uniform throughout the circuit, we may write

$$H_{fxd} d = -\Phi_0 R_{fxc}, \quad \ldots\ldots \quad (3)$$

whilst

$$\Phi_0 = A_{fxd} B_{fxd},$$

the indices fxd and fxc standing for ferroxdure and ferroxcube respectively; for a completely filled gap $A_{fxd} = A_s$. From the two equations it follows that

$$\frac{B_{fxd}}{-H_{fxd}} = \frac{d/A_{fxd}}{R_{fxc}},$$

or, according to (2),

$$\frac{B_{fxd}}{-\mu_0 H_{fxd}} = \frac{R_s}{R_{fxc}}, \quad \ldots\ldots \quad (4)$$

which latter ratio, by assumption, is known.

In the idealized B-H curve of ferroxdure (*fig. 9*), equation (4) is represented by the straight line *0-1* (the working line). This line should be independent of the choice of d and A_s as long the ratio d/A_s is kept constant.

The intersection of the working line and the B-H curve defines the working point of the ferroxdure

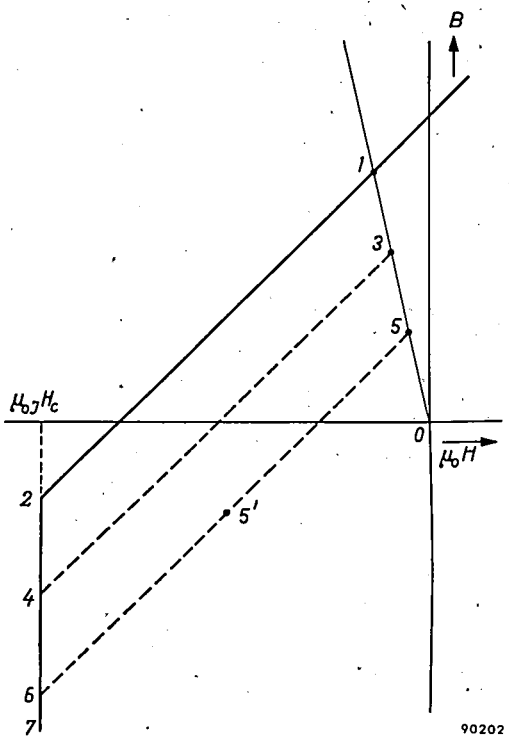


Fig. 9. *1-2-7* is part of the B-$\mu_0 H$ curve of strongly magnetized ferroxdure, *3-4-7* the same for a more weakly magnetized ferroxdure, and *5-6-7* for a still more weakly magnetized ferroxdure. *0-1* working line of a ferroxdure magnet incorporated in a magnetic circuit. Working point *1* is impracticable (safety factor less than *1*), *3* represents the critical case (safety factor = 1), and *5* permits a safety factor = 2.

when the transformer is currentless. The three parallel lines in fig. 9 are applicable to three different transformer sizes. Let us first choose d and A_s such that completely magnetized ferroxdure is necessary to produce the desired flux Φ_0 in the currentless transformer. This is the flux that brings out about an induction of 0.3 Wb/m² in the ferroxcube core and is represented by the point R in fig. 5. The working point would be situated at *1* (fig. 9). It is evident that if the flux is reversed, a point *2* would be passed, the latter being situated less below the horizontal axis than *1* lies above it. This case, therefore, cannot be used, and we have to increase A_s and d and use less strongly magnetized ferroxdure. If this is continued until the B value corresponding to point *3* is just adequate to produce the desired flux Φ_0, we arrive at the critical condition (safety factor 1): if the flux is reversed point *4* is reached, which lies equally far from the axis as *3*.

In practice, however, a certain safety margin is required, which is provided by choosing point *5* as working point. If the flux is reversed, *5'* becomes the working point (the ordinate of *5* is B_0, that of *5'* is $-B_0$). The reverse flux is now allowed to assume such a value that the point *6* with ordinate $-3B_0$ is reached, the safety factor being the ratio of the length *5-6* to the length *5-5'*, which is 2.

The procedure for determining the dimensions d and A_s is now a simple one. The pre-magnetizing flux needed to produce the desired pre-induction B_0 (= 0.3 Wb/m² in fig. 5) in the core is $\Phi_0 = A_k B_0$. Let B_w be the ordinate of the working point corresponding to a sufficiently large safety factor (such as point *5* in fig. 9), then the required cross-section A_{fxd} for the ferroxdure is given by:

$$A_{fxd} = \Phi_0 / B_w = A_k B_0 / B_w .$$

From this cross-section, by means of eq. (2), we find the necessary thickness d of the ferroxdure slab (this being also the width of the gap).

Partly filled gap

On the foregoing theoretical considerations we have arrived at a workable design for a transformer with a gap completely filled with incompletely magnetized ferroxdure (fig. 8*a*). The question now arises whether it is also possible (and perhaps better) to use a gap partly filled (fig. 8*b*) with more strongly magnetized ferroxdure. We find that the latter case can be easily related to the foregoing as follows.

If we reduce the cross-section of the filling in a given gap ($A_{fxd} < A_s$, fig. 8*b*), then obviously the magnetization of the ferroxdure must be raised to such an extent that the flux Φ_0 through the core remains the same. This means that the magnetic potential difference between the boundary planes of the gap, $\Phi_0 R_{fxc}$, must retain the same value. It follows from (3), therefore, that H_{fxd} must not change either. By analogy, we can also demonstrate that under conditions of reversed flux the value of H_{fxd} must remain the same. Let *a-b* in *fig. 10* be the earlier mentioned working line for a completely filled gap, then the working line for a gap appropriately filled with completely magnetized ferroxdure will be *c-d*, which lies parallel to and vertically above *a-b*. The corresponding change in induction ΔB is hence equally large in either case, since an equal flux variation $2\Phi_0$ is evenly distributed throughout the

gap, irrespective of what portion of the gap is taken up by air and what by ferroxdure (both being here assumed to have $\mu_r = 1$); in either case the change ΔB will be $2\Phi_0/A_s$.

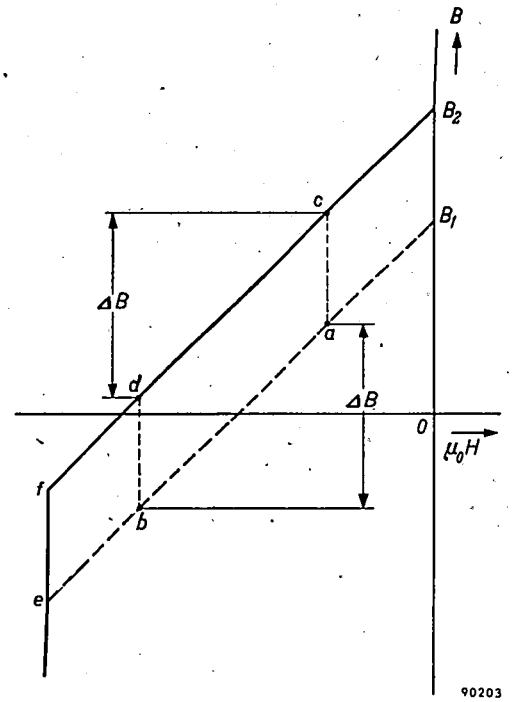


Fig. 10. *a-b* working line of a ferroxdure magnet completely filling the gap of the pulse transformer; remanence B_1. If the gap is partly filled with more strongly magnetized ferroxdure (remanence B_2), the working line becomes *c-d*, situated vertically above *a-b* and along the characteristic B_2-*f*, which runs parallel with B_1-*e*. The change in induction ΔB as well as the safety factor is the same in either case (at least in these idealized curves).

We further notice that the maximum range through which B can travel without causing permanent demagnetization likewise remains unaltered: accordingly the safety factor ($ae/ab = cf/cd$ in fig. 10) is also unaltered. The two cases depicted here are completely equivalent. Later we shall see, however, when the actual *B-H* curve is substituted for the idealized straight-line representation used up to now, that designs using the completely filled gap are to be preferred.

The question now arises to what extent the gap has to be filled with more strongly magnetized ferroxdure; in other words what must be the ratio $a = A_{fxd}/A_s$?

The equivalence of the two cases considered in fig. 10 is a result of the fact that the only relevant factor in this question (the ratio R_s/R_{fxc} being supposed constant) is the short-circuit flux of the ferroxdure, $A_{fxd}B_{rem}$, irrespective of the manner in which the gap is filled. Therefore, when a core with completely filled gap is modified into one with a gap partly filled with more strongly magnetized material, it is essential to keep $A_{fxd}B_{rem} = aA_sB_{rem}$ constant.

When changing over from *a-b* to *c-d* in fig. 10, we must choose $a = B_1/B_2$ (A_s not being changed), and for intermediate cases, where the remanence is B_{rem}, the working line will as a rule lie between the vertical dotted lines, provided that a and B_{rem} are so chosen that $aB_{rem} = B_1$.

This concludes our general theoretical treatment of the subject. We shall now consider the actual *B-H* curve and find out to what extent our conclusions have to be modified.

Actual *B-H* curve of ferroxdure

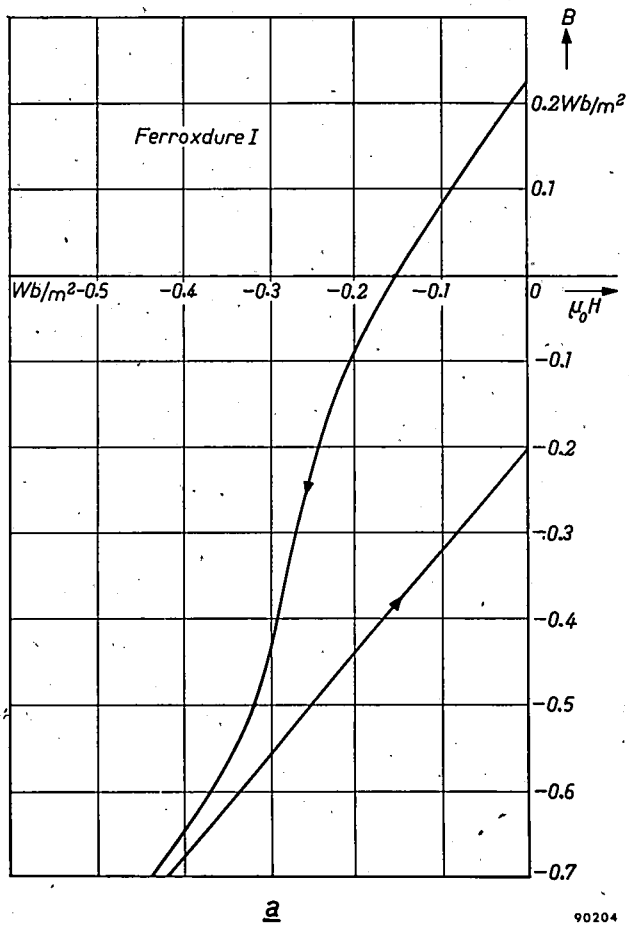
Fig. 11a shows the B-μ_0H curve measured on ferroxdure I. The part we are primarily concerned with is represented on an enlarged scale in fig. 11*b*, together with the reversible characteristics for $B_{rem} = 0.20$ and 0.12 Wb/m^2.

In the following we shall retain a constant ratio R_s/R_{fxc}, which we shall assume to be unity ($R_s = R_{fxc} = \frac{1}{2}R$), which was found to be a convenient value in practice. For a completely filled gap ($a = 1$), according to eq. (4), the working line in the B-μ_0H diagram will run at an angle of 45°.

Let us examine what possibilities these two conditions of magnetization may offer. With $B_{rem} = 0.20$ Wb/m^2, we find the working point *1*, where $B_{fxd} = 0.08$ Wb/m^2. In order to produce the desired pre-induction $B_0 = 0.30$ Wb/m^2 in the core, we must therefore choose $A_s = (0.30/0.08)A_k = 3.8\ A_k$. We find, however, from fig. 11*b* that the intersection *2* with the *B-H* curve lies less below the axis than *1* lies above it, so that reversing the polarity is not permissible. This is also manifest from the fact that the safety factor is less than 1, the value for $(\Delta B)_{max}$ in ferroxdure reading 0.13 Wb/m^2 in the diagram, so that the safety factor is as low as $0.13/(2\times 0.08) \approx 0.8$.

The weaker pre-magnetization $B_{rem} = 0.12$ Wb/m^2 gives a far better result. The working line becomes *3-3'* in fig. 11*b*, with the value $B_{fxd} = 0.05$ Wb/m^2 in point *3*, so that A_s must become $(0.30/0.05)A_k = 6\ A_k$. The required quantity of ferroxdure, therefore, is $(6/3.8)^2 = 2.5$ times that in the previous example (both the cross-section A_s and the thickness d increase by a factor of 6/3.8, their ratio being constant according to (2)). This larger amount of material results in a remarkable improvement of the safety factor. This factor, being the ratio of the length of *3-4* to that of *3-3'*, now becomes 2.5, which may be considered adequate.

We now have a workable design with weakly magnetized ferroxdure filling the entire air-gap.



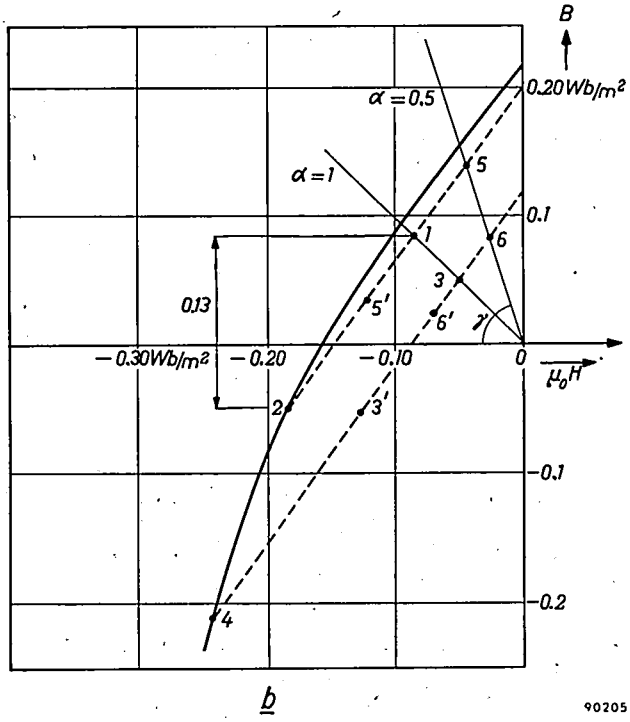


Fig. 11. *a*) B-$\mu_0 H$ curve measured on ferroxdure I. *b*) Part of (*a*) enlarged, with the reversible characteristics for $B_{rem} = 0.12$ and 0.20 Wb/m², and the working lines for $\alpha = 1$ and 0.5. Working point *1* cannot be used (safety factor less than 1). The curve *3-3'* corresponds to working point *3*, *5-5'* corresponds to *5*, and *6-6'* to *6*.

From this, we might now change over to a partial filling with stronger magnetic material, in the manner described earlier by shifting the working-line *3-3'* vertically upward, whilst appropriately varying the value of α.

We may, however, also approach the matter differently again, viz. by assigning a fixed value (< 1) to α and then investigating the possibilities thus offered. We shall do so for $\alpha = 0.5$ in combination with the values $B_{rem} = 0.12$ and 0.20 Wb/m².

Let us start by finding the working line for the currentless condition. For a half-filled gap we must replace R_s in the numerator of (4) by the magnetic resistance of the half gap, viz. $2R_s$ ($= 2R_{fxc}$); the denominator contains the magnetic equivalent resistance of the ferroxcube core shunted by the empty half of the gap: R_{fxc} in parallel with $2R_s = R_{fxc}$ in parallel with $2R_{fxc} = \frac{2}{3} R_{fxc}$. The slope of the working line for $\alpha = 0.5$ (see fig. 11*b*) is therefore $\tan \gamma = 2R_{fxc}/(\frac{2}{3}R_{fxc}) = 3$. This working line intersects the two reversible characteristics in the points *5* and *6*. In order to evaluate the induction B_0 in the ferroxcube core, the flux through the empty half of the gap (induction B_l) must be subtracted from the flux produced by the ferroxdure. In the empty half H equals H_{fxd}. We may therefore write:

$$A_k B_0 = \tfrac{1}{2} A_s (B_{fxd} + B_l) = \tfrac{1}{2} A_s (B_{fxd} + \mu_0 H_{fxd}), \quad (5)$$

in which B_l and H_{fxd} are negative and have values that can be read from fig. 11*b*. This permits us to evaluate the ratio A_s/A_k necessary to produce the pre-induction $B_0 = 0.30$ Wb/m². Thus we find the values $A_s/A_k = 6.7$ for point *5* and $A_s/A_k = 12$ for point *6*. The working points for the reverse flux — which we must know for determining the safety factor — can be found by substituting the above values of A_s/A_k and B_0 in (5). For the two cases considered here we arrive at the conditions

$$B + \mu_0 H = -0.09 \text{ and } -0.05 \text{ Wb/m}^2 \text{ respectively,}$$

corresponding to the points *5'* and *6'* in fig. 11*b*. The safety factor can now be read from the diagram: it amounts to $(5\text{-}2)/(5\text{-}5') = 1.8$ for the material with $B_{rem} = 0.20$ Wb/m², and to $(6\text{-}4)/(6\text{-}6') = 5.0$ for that with $B_{rem} = 0.12$ Wb/m².

In *Table I*, finally, we can summarize the cases dealt with in this section. For each of the four cases considered, are given the values of A_s/A_k, the short-circuit flux $\alpha B_{rem} A_s$ of the ferroxdure divided by A_k, the required volume of ferroxdure (as a percentage), and the safety factor.

Table I. Characteristic quantities relating to a pulse transformer core, pre-magnetized with ferroxdure I, for values of $B_{rem} = 0.12$ and 0.20 Wb/m² and for a completely-filled gap ($\alpha = 1$) and for a half-filled gap ($\alpha = 0.5$), R_s being equal to R_{fxc}.

	$\alpha = 1$		$\alpha = 0.5$	
B_{rem} (Wb/m²)	0.12	0.20	0.12	0.20
A_s/A_k	6.0	3.8	12	6.7
$\alpha B_{rem} A_s/A_k$ (Wb/m²)	0.72	0.74	0.72	0.67
Volume of ferroxdure (%)	260	100	530	160
Safety factor	2.5	0.8	5.0	1.8

Row 1 of the table shows that A_s/A_k is always fairly large, so that the ferroxcube surfaces in the air-gap must be substantially larger than the core diameter within the coil. This leads, as we shall see presently, to constructions different from those shown in fig. 8.

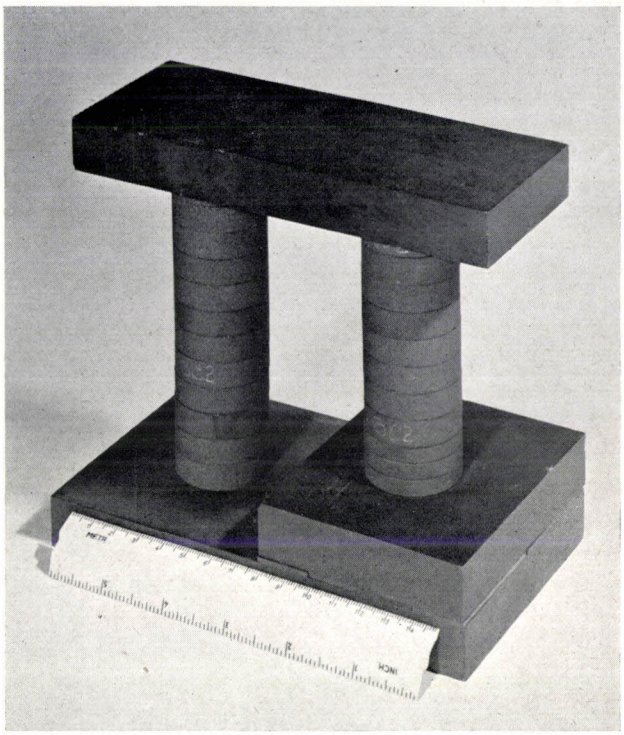
On page 34, on the strength of the idealized characteristic and on the assumption that $\mu_r = 1$ for ferroxdure, we have said that the short-circuit flux $\alpha A_s B_{rem}$ of the ferroxdure must be held constant. The actual values of this flux (divided by the constant cross-section A_k) as given in row 2 of the table, are indeed found to be fairly constant for the actual characteristic as well.

Rows 3 and 4 show the surprising result that the safety factor in the region considered is roughly proportional to the ferroxdure volume, and independent of its shape and state of magnetization (α and B_{rem}). It will, therefore, always be preferable to use a completely filled gap, since the ferroxcube yoke on either side of the gap can then have the smallest diameter, whereas on the other hand the saving in ferroxdure by using a partially filled gap is not worth mentioning. Of the four cases shown in the table that of column 1 is preferable for this reason. If a larger safety factor, e.g. 5, is required, it should be possible with $\alpha = 1$ to arrive at a better design than that of column 3. Owing to the inadequate safety factor, the case of column 2 cannot be used, whilst that of column 4 has a somewhat low safety factor and is, moreover, in view of $\alpha < 1$, by no means the best solution.

Example of an actual design

Finally, as an example we shall describe a pulse transformer operating in a laboratory set-up for testing magnetrons. The primary demands for this transformer were that it should produce a secondary pulse of the appropriate shape and that it should retain its pre-magnetization under all conditions.

Fig. 12 is a photograph of the core and the essential data are given in the caption. The completely filled gap is 2 mm wide and has an area $A_s = 8 \times 10 = 80$ cm², i.e. 8 × larger than the cross-section of the cylindrical cores. This large ratio A_s/A_k, together with the wish to use ferroxcube blocks of standardized dimensions, have led to the shape depicted here.



89953

Fig. 12. Core of the pulse transformer of a testing installation for magnetrons. Principal dimensions: base 100 × 170 mm, height 162 mm. Further principle data:

	Primary	Secondary
Peak voltage	7 kV	28 kV
Peak current	140 A	35 A
Impedance	50 Ω	800 Ω
Number of turns	16	66
Pulse duration	1 μsec	
Rise time	0.14 μsec	
Cross-section A_k within coil	10 cm²	
Area A_s of gap	80 cm²	
Width d of gap	2 mm	
Safety factor	approx. 3.	

The two vertical columns consist of ferroxcube 3C2 discs of 10 mm thickness (thicker discs are difficult to manufacture without cracks) and 36 mm diameter. The yokes are formed by rectangular blocks of ferroxcube 4B2, which can easily be manufactured in large pieces; if necessary, the somewhat smaller permeability can be compensated by giving the yokes a larger cross-section. Pre-magnetization is effected by eight 2 × 20 × 50 mm slabs of ferroxdure I, magnetized in the direction of their thickness. These are interposed between two large slabs

of ferroxcube, the upper one serving to concentrate the flux into the much smaller cross-section of the cylindrical part of the core. The safety factor is about 3. Each column is wound with a primary and a secondary winding, the primary windings being connected in parallel and the secondary windings in series. The whole assembly, clamped together by rods of insulating material, is immersed in oil. It should be noted that for a laboratory installation the dimensions are of secondary importance, so that no attempt has been made to minimize the dimensions of the transformer.

Summary. With transformers in which one of the windings carries unidirectional pulse-shaped currents, pre-magnetization of the core forms a means to arrive at a more favourable loading of the core. As an example the pulse transformer of a radar transmitter is discussed. To give the secondary pulse the appropriate shape, the transformer must have a considerable self-inductance, whilst the wire length of the turns must be kept short. Pre-magnetization of the core allows a more favourable compromise between these two conflicting demands. This can be effectively realized by incorporating in the magnetic circuit composed of ferroxcube blocks, a slab of magnetized ferroxdure. If it is arranged that the pre-magnetization brings about an induction of -0.3 Wb/m^2 in the ferroxcube within the coil, and that when pulses are applied to the transformer, this induction can fluctuate between -0.3 and $+0.3$ Wb/m^2, then the relative permeability of the core can reach a value as high as about 300.

Dimensions should be judiciously chosen to preclude the possibility of permanent demagnetization of the ferroxdure by the pulses. The theory of magnetic circuits incorporating a permanent ferroxdure magnet and subjected to a reversing flux is treated with the aid of an idealized straight-line magnetization curve, and subsequently verified with the aid of the actual curve. Finally a practical example of such a transformer is described.

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.

2415: H. O. Huisman, A. Smit, P. H. van Leeuwen and J. H. van Rij: Investigations in the vitamin A series, III. Rearrangement of the retro-system to the normal system of conjugated double bonds in the vitamin A series (Rec. Trav. chim. Pays-Bas **75**, 977-1006, 1956, No. 7).

A new synthesis of vitamin A essentially based on the almost quantitative rearrangement of the retro-system to the normal system of conjugated double bonds is described. This rearrangement is carried out by converting the retro-C_{15}-acid and the retro-C_{20}-acid, by means of phosphorus trichloride in suitable solvents, into the corresponding retro-acid chlorides, which thereupon rearrange smoothly to the acid chlorides, with the normal system of conjugated double bonds, β-ionylidene acetic acid chloride and vitamin A acid chloride respectively. The acid chlorides are reduced with $LiAlH_4$ to the corresponding alcohols β-ionylidene ethanol and vitamin A. The physical and chemical properties of the intermediates and geometric isomers obtained in the various reaction steps are described.

2416: F. A. Kröger, G. Diemer and H. A. Klasens: Nature of an ohmic metal-semiconductor contact (Phys. Rev. **103**, 279, 1956, No. 2).

Models for an ohmic contact between a metal and a high-ohmic *n*-type semiconductor like CdS as proposed by Smith and Butler are not quite satisfactory with regard to experimental evidence. An alternative model is proposed, according to which a thin layer of the semiconductor adjacent to the electrode is strongly *n*-type, e.g. due to diffusion of trivalent metal ions from the electrode into the CdS or to the bombardment of the CdS surface, which is often used to clean the surface before applying the electrodes.

2417: P. A. Neeteson: The vacuum tube as a network component in pulse circuits (T. Ned. Radiogenootschap **21**, 171-185, 1956, No. 4).

In this article a survey and some illustrating examples are given of methods by which large-signal behaviour of vacuum tubes can be determined. Contrary to small-signal application, where the tube is operated by relatively small deviations around a fixed operating point in the conducting region, in large-signal operation the tube is rapidly brought from the cut-off into the conducting region and *vice versa*. This is essentially a switching action. Therefore, it is necessary to give some preliminary considerations on switch operation in networks. The method has proved to be useful in analyzing electronic pulse circuits. See also the book by the same author, announced in Philips tech. Rev. **18**, 284, 1956/57.

2418: A. van Wieringen and N. Warmoltz: On the permeation of hydrogen and helium in single crystal silicon and germanium at elevated temperatures (Physica **22**, 849-865, 1956, No. 10).

A mass spectrometer examination of the permeability of the elements silicon and germanium to the gases hydrogen and helium has been carried out in the temperature range 967-1207 °C for silicon and 766-930 °C for germanium. Using certain crystal growing and cutting techniques, two kinds of diffusion cells were made by each of which it was possible to determine both diffusion coefficient and solubility as well as the activation energies of diffusion and solution from non-steady-state permeation measurements. No permeation of the gases neon, argon and nitrogen could be detected. It seems that hydrogen in silicon can occur in a readily-diffusible form and also in a state in which it is less mobile. The readily-diffusible form consists of atoms or protons.

2419: N. W. H. Addink: Spectrochemical analysis by means of the D.C. carbon arc (Appl. Spectroscopy **10**, 128-137, 1956, No. 3).

A complete description of the constant temperature arc method of quantitative analysis which has been developed in Eindhoven is given, with tables of the empirically determined K-values reported so that they can be checked in other laboratories. The method consists of completely volatilizing 5 mg of a powdered sample in a shallow anodic crater of a carbon arc, with the addition of materials to modify the rate of volatilization if required. The line intensities are calibrated and corrections are made by comparison with selected Fe lines, originating from a "standard light source" so as to get comparable analytical results; the calculations required are illustrated by several examples, which indicate the relative accuracy of the method to be approximately 10%.

2420: K. Compaan and Y. Haven: Correlation factors for diffusion in solids (Trans. Faraday Soc. **52**, 786-801, June 1956, No. 6).

The relation of Einstein $D = BkT/e$, relating the diffusivity and mobility of particles (ions or atoms), must be modified for the case of diffusion in solids, if a vacancy mechanism holds, because there will be a correlation between successive steps of a particle, even if the steps of the vacancies themselves are uncorrelated. Certain symmetry conditions being fulfilled, the relation of Einstein must be modified by a correlation factor

$$f = (1 + \overline{\cos \vartheta_{i,i+1}})/(1 - \overline{\cos \vartheta_{i,i+1}}), \qquad (1)$$

where $\vartheta_{i,i+1}$ is the angle between two successive steps of a particle. The diffusion problem can be translated into the theory of electrical networks and with the help of measurements in a resistor network the value of $\overline{\cos \vartheta_{i,i+1}}$ is easily evaluated. This has been done for several types of lattices. Other cases, where eq. (1) does not hold, because of lack of symmetry, are treated by similar methods. Appropriate substitutions having been made, the diffusion of associated pairs has been treated as the diffusion of a single particle in a so-called transposed lattice.

2421: C. M. van der Burgt: Les transducteurs piézomagnétiques à noyau massif de ferrite (Commun. Congrès international sur les traitements par les ultra-sons, Marseille 23-28 May 1955; published 1956).

Interim report on some of the results of an investigation into the use of ferroxcube materials for piezomagnetic vibrators. A fuller account has meanwhile appeared in this Review: Philips tech. Rev. **18**, 285-298, 1956/57, (No. 10).

2422: R. Vermeulen: Stereo reverberation (IRE Transactions on Audio, **AU-4**, 98-105, 1956, No. 4).

Reproduction of the article published in Philips tech. Rev. **17**, 258-266, 1955/56.

2423: H. J. Oskam: High-frequency gas-discharge breakdown in neon-argon mixtures (J. appl. Phys. **27**, 848-853, 1956, No. 8).

Breakdown electric fields in a waveguide at a frequency of 9500 Mc/s are presented for a number of neon-argon mixtures at various pressures. The argon percentage is found to have a large influence on the breakdown electric field, just as Penning found for the D.C. discharge. For each argon concentration, only one minimum is found in the curves giving breakdown field as a function of pressure. This result contrasts with the two minima found in the D.C. discharge for some neon-argon mixtures. This can be explained by the difference between the efficiency of energy transfer from the electric field to the electrons in the D.C. case and the high-frequency case. The relation between the concentration of argon and the breakdown fields is discussed. The influence of electric field distortions and the standing-wave ratio in the waveguide are investigated.

2424: W. L. Wanmaker: Contribution à la chimie des halophosphates de calcium (J. Phys. Radium **17**, 636-640, 1956, No. 8-9).

Detailed study of the secondary reactions during

the preparation of calcium halophosphates activated by Sb^{3+} or by Sb^{3+} and Mn^{2+}, in order to suppress the formation of Sb^{5+} or Mn^{3+} ions, or of free Sb_2O_3 and Mn_3O_4, the effect of which is to lower the quantum efficiency.

2425: J. L. Ouweltjes: Quelques considérations sur la transmission d'énergie dans les halophosphates (J. Phys. Radium **17**, 641-644, 1956, No. 8-9).

Measurements have been made of the brightness of halophosphates activated by Sb and Mn, and of the killer effect of iron. The energy transfers between activator, sensitizer and killer ions are discussed. Possibly the greatest part of the killer effect of iron is a mere absorption.

2426: W. Ch. van Geel: Sur la luminescence de couches d'oxydes formées par oxydation électrolytique (J. Phys. Radium **17**, 714-717, 1956, No. 8-9).

Measurements of the luminescence L produced during the formation of layers obtained by anodic oxidation give $L = aI(e^{bd} - 1)$; I current density, d thickness of the layer. When A.C. tension is applied, light flashes are observed at each change of polarity.

2427: P. Zalm: Sur l'électroluminescence du sulfure de zinc (effet Destriau) (J. Phys. Radium **17**, 777-782, 1956, No. 8-9).

A discussion is given of the voltage and temperature dependence of the emittance of an electroluminescent phosphor. The emittance (H) — voltage (V) relation is given by $H = H_0 \exp(-b/V^{1/2})$. The variation of the emittance with temperature depends on the relation between the local field in the phosphor particles and the applied field, the temperature dependence being described by the variation of b with temperature.

2428: J. H. Stuy: Studies on the mechanism of radiation inactivation of micro-organisms, II. Photoreactivation of some bacilli and of the spores of two Bacillus cereus strains (Biochim. biophys. Acta **22**, 238-240, 1956, No. 2).

In order to find out whether bacilli in general could be reactivated by light after ultraviolet inactivation, several different species have been investigated. About half of them did not show any photoreactivation (PHR) while some of them showed a moderate PHR. Only two B. cereus strains could be photoreactivated very easily. The conclusion is drawn that PHR is not so generally occurring among the bacilli, but under entirely different conditions there might be some reactivation. The spores of the B. cereus strains mentioned above did not show PHR. This was not due to their lack of water, since "germinated spores" behaved similarly in this respect. Germination of the spores was carried out in a synthetic medium which permitted a transformation of the spores into a following stage only. During this transformation solids are exchanged for water from the medium, causing a considerable increase of the water content.

2429: J. H. Stuy: Studies on the mechanism of radiation inactivation of micro-organisms, III. Inactivation of germinating spores of Bacillus cereus (Biochim. biophys. Acta **22**, 241-246, 1956, No. 2).

Upon incubation in a germination medium, resting spores of Bacillus cereus lost their resistance against heat, X-rays and ultraviolet radiation. These phenomena were studied in detail using a synthetic germination medium which permitted a transformation of the spores into a following stage only. The results showed that the loss in resistance against heating at 70 °C for 10 minutes was very rapid. The loss of resistance against X-rays followed roughly the same rate. The UV-resistance of the spores, however, increased considerably during the first minute of incubation; thereafter it rapidly fell off. Apart from the UV-increase in the first minute, the simultaneous loss of the three resistances studied suggests that this loss is due to one mechanism. In this respect the water uptake by the spores is considered.

2430: H. B. Haanstra: Einige Bemerkungen über die Polystyrol-SiO-Abdruckmethode (Rev. universelle Mines **99**, 481-485, 1956, No. 10). (Remarks on the polystyrol-SiO replica technique; in German.)

The examination of electron micrographs obtained using the polystyrene-SiO technique shows that the thickness of the SiO layer is not uniform. Using this idea, the author explains that certain zones of the replica which were not perpendicular to the direction of the beam, will appear much clearer than others in transmission.

2431: J. B. de Boer and W. Morass: Berechnung der Sehweite aus der Lichtverteilung von Automobilscheinwerfern (Lichttechnik **8**, 433-437, 1956, No. 10). (Calculation of the visibility range from the light distribution of car headlamps; in German.)

V. J. Jehu has derived a partly experimental,

partly theoretical method for calculating the visibility range in the light of car headlamps of which the light distribution is known. Allowance is made for dazzle by oncoming cars. The results, however, are not always in agreement with observation. The authors have therefore made extensive new measurements of visibility under road conditions. With the help of the results the visibility can be derived graphically. Comparisons between the quality of car headlamps can then be made very simply, knowing their light distribution.

R 310: C. M. van der Burgt: Controlled crystal anisotropy and controlled temperature dependence of the permeability and elasticity of various cobalt-substituted ferrites (Philips Res. Rep. **12**, 97-122, 1957, No. 2).

Dynamic elasticity at remanence, piezomagnetic coupling at remanence, and initial permeability of polycrystalline toroids of various cobalt-substituted ferrites are determined in the temperature range from -196 °C to $+120$ °C (and higher). The compositions are represented approximately by $(M_{1-y}Zn_y)_{1-x}Co_xFe_2O_4$, where M stands for one of the divalent metal ions ($Li_{\frac{1}{2}}^{1+}Fe_{\frac{1}{2}}^{3+}$), Ni^{2+}, or Mn^{2+}. The first-order magnetocrystalline anisotropy constant K_1 of cobalt ferrite is known to be large and positive (2 to 3×10^5 J/m^3). Consequently the incorporation of a small amount of cobalt ferrite in solid solution in other ferrites that have $K_1 < 0$ (ferrous, nickel, manganese ferrites, and presumably lithium ferrites) leads to a compensation of crystal anisotropy at a transition temperature T_0 depending on the amount of cobalt ferrite. At room temperature the compensating effect of cobalt substitutions in ferrous ferrite is nearly 4 times as strong as expected from simple linear interpolation. On the other hand cobalt substitutions in manganese ferrite appear to be less effective than expected, the compensating amount of cobalt being about 4 times that expected. Nevertheless the curves T_0 vs. x for all systems of cobalt-substituted mixed ferrites have a similar shape. The fact that the crystal anisotropy passes through zero at a temperature T_0 implies that around this temperature only strain anisotropy and pore-shape anisotropy remain, so that there exists a small temperature range where the substance is magnetically and magnetoelastically soft. The resulting peaks in the permeability and compliance, together with the overall increase of these quantities with temperature, lead to temperature ranges some-where above T_0 where the permeability and the elasticity are substantially temperature-independent.

R 311: K. S. Knol: A thermal noise standard for microwaves (Philips Res. Rep. **12**, 123-126, 1957, No. 2).

A thermal noise source for the 3-cm waveband, consisting of a heated platinum waveguide terminated with a ceramic wedge, is described. The temperature of the wedge is fixed at the melting-point of gold. With this noise source as a standard, the noise temperature of the Philips noise source K50A is determined to be 21700 °K, with an accuracy of about 5 per cent.

R 312: J. van den Boomgaard and K. Schol: The P-T-x phase diagrams of the systems In-As, Ga-As and In-P (Philips Res. Rep. **12**, 127-140, 1957, No. 2).

For the compounds In-As, Ga-As and In-P the phase relations solid-liquid-vapour have been determined. In-As has a maximum melting-point of 943 ± 3 °C at an arsenic pressure of 0.33 atm. Ga-As has a maximum melting-point of 1237 ± 3 °C at an arsenic pressure of 0.9 atm. For In-P the maximum melting-point is estimated to lie at 1062 ± 7° C at a phosphorus pressure of approximately 60 atm.

R 313: J. L. H. Jonker and Z. van Gelder: The internal resistance of a radio-frequency pentode (Philips Res. Rep. **12**, 141-175, 1957, No. 2).

Besides the direct electrostatic influence of the anode potential upon the cathode current, which is small in radio-frequency pentodes, the authors investigate the effects causing the current distribution between screen grid and anode due to reflection of electrons by the suppressor-grid wires and by the anode (secondary emission). The extra space charge in the cathode space, due to these reflected electrons, is taken into account. The reflected electrons are also distributed between anode and screen grid, as a result of the same effects. A convergent series therefore originates for the anode current, the derivative d_aI/dV_a of which gives the reciprocal value of the internal resistance. To obtain a high value of the internal resistance, the reflection coefficient of the suppressor grid must be small, which can be obtained by taking a very small value for the ratio between the wire diameter and the pitch of this grid. The effective potential of the suppressor grid, however, must be kept low. The calculated values of the internal resistance of radio-frequency pentodes are about 40% higher than the measured ones, which, in view of the large number of effects playing a part, must be considered as a satisfactory result.

VOLUME 19, 1957/58, No. 2 pp. 41-72 Published 24th August 1957

Philips Technical Review

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

AUDIO AMPLIFIERS WITH SINGLE-ENDED PUSH-PULL OUTPUT

by J. RODRIGUES de MIRANDA.

621.375.2.029.3

For some time now radio receivers have been available in which the loudspeaker is directly driven by the output valves, without the use of an output transformer. This means the omission of a component that, apart from adding to the losses and distortion, is quite expensive even to satisfy only moderate demands. The phase shift it introduces, moreover, prevents feedback being applied to its full advantage. This article deals with a new transformerless circuit for the last two stages of audio-frequency amplifiers which differs appreciably from the conventional one and can satisfy far higher demands. No expensive components are required to replace the transformer.

Music reproduction in the home of gramophone records, tape recordings or F.M. transmissions, may nowadays reach a very high quality indeed. If the excellent properties of such "music sources" are to be done full justice, the reproduction installation must come up to very high standards. Particularly in the last few years, therefore, there has been a general trend towards improvement of the reproduction system. This may consist of one or more loudspeakers and an audio-frequency amplifier, connected to a pick-up, the play-back head of a tape recorder, or to the R.F. and I.F. stages of an F.M. receiver.

Confining ourselves to the audio amplifier and the loudspeaker alone, there are three aspects to be considered, viz. the electronic circuitry, the loudspeaker design, and the acoustics.

The last two aspects were dealt with at some length in a recent article in this Review on a high-fidelity loudspeaker installation [1]). The present article is mainly concerned with the electronic aspects, i.e. with circuit development. As we shall see presently, this work has been influenced by certain innovations in the loudspeaker field.

The demands made on audio-frequency amplifiers involve the output power, frequency response, non-linear distortion and output resistance.

As regards frequency-response characteristics, it is not sufficient if merely the *amplitude* characteristic is flat within the audible range (20-18 000 c/s), since then the *phase* characteristic would possess the appropriate form only within a far narrower range; by the "appropriate" form is meant that the phase angle is proportional to the frequency (the proportionality factor may also be zero). With a poor phase characteristic a sudden burst of music, sharp consonants, etc. cannot be reproduced without distortion [2]). A good phase characteristic can be obtained if the amplitude characteristic remains flat far beyond the audible range, e.g. up to 40 kc/s. However, it is usually considered desirable to be able to vary the amplitude characteristic at either end of the audio band, i.e. have a tone control — preferably independent — of both the high and low notes.

Non-linear distortion in an amplifier is due to the non-linear characteristics of valves and magnetic materials. It gives rise to harmonics, which alter the timbre of the sound, and, worse, when two or more tones are produced simultaneously, to non-harmonic overtones (intermodulation). Non-linear distortion, rapidly increasing with amplitude above a certain signal strength, puts a definite limit to the maximum output power at which the amplifier can be satisfactorily operated.

[1]) G. J. Bleeksma and J. J. Schurink, A loudspeaker installation for high-fidelity reproduction in the home, Philips tech. Rev. **18**, 304-315, 1956/57 (No. 10).

[2]) See e.g. J. Haantjes, Judging an amplifier by means of the transient characteristic, Philips tech. Rev. **6**, 193-201, 1941.

Finally, the output resistance of the amplifier should be low, in order to provide a substantial damping effect on the loudspeaker, the latter having a mechanical resonance in the region of the very low notes [3]).

The designer of the amplifier has to try to satisfy these demands on characteristics, distortion and output resistance, whilst keeping the costs as low as possible. Considerations of price may carry considerable weight where mass production, e.g. of radio sets, is concerned. A low cost price is closely related to a high efficiency, for a poor efficiency involves a larger power pack and unnecessarily high-power output valves.

As output valve, the pentode is generally preferred to the triode, the former having a higher efficiency and requiring a smaller input signal. With a pentode, maximum efficiency (theoretically 50%) is obtained if the load resistance R_0 equals the D.C. resistance R_{ak} of the valve ($R_{ak} = V_{a(w)}/I_{a(w)}$, $V_{a(w)}$ and $I_{a(w)}$ being the anode voltage and the anode current at the working point; R_{ak} should not be confused with the (much higher) internal resistance $R_i = \partial V_a/\partial I_a$ at constant control-grid voltage). For most output pentodes R_{ak} has a value of several thousands of ohms.

A moving-coil loudspeaker of conventional design has a speech-coil of relatively few turns of fairly thick wire. Its impedance (virtually a resistance) is accordingly low, e.g. 7 Ω. This low impedance is matched to the far higher value of R_{ak} via a step-down output transformer. This component requires a great deal of care and material in its construction. Its stray self-inductance forms together with stray capacitances one or more oscillatory circuits, which impose an upper limit to the frequency range that can be reproduced. This cut-off frequency is the higher as the stray field and the winding capacitance are smaller, and therefore as the number of turns is smaller. One essential requirement for a good reproduction of the *low* notes, however, is a large primary self-inductance, which means a large number of turns. Another unfavourable factor is the fact that the anode direct current pre-magnetizes the core of the transformer.

The latter difficulty is avoided if a push-pull output stage is employed; D.C. magnetization of the core is then absent and a more favourable compromise between large primary self-inductance and small stray field and capacitance can be reached. A push-pull arrangement has furthermore the well-known advantage of producing less distortion: if the circuit is symmetrical, no even harmonics occur in the output signal. The complications incurred by using a second output valve (with the necessary phase-inverting element for getting two equal driving voltages in phase opposition), make the circuit more expensive, but for better-quality radio receivers and amplifiers the above advantages of push-pull operation outweigh the difference in price.

Apart from limiting the frequency range at both ends, the output transformer has the drawback of causing a loss of output power. Its efficiency is often as low as 50%, and seldom better than 80%, dependent upon the amount of material (and hence the cost) spent on it. The same applies to the distortion it introduces as a result of the non-linear *B-H*-curve of the iron core; here too, improvement can be reached only at the expenditure of more material.

Most serious of all, perhaps, is the fact that the output transformer prevents *negative feedback* being applied to its full advantage. With the aid of negative feedback, as is well known, the amplitude characteristic can be improved, the non-linear distortion reduced, and the output resistance lowered. If the circuit contains a phase-shifting element, however, such as an output transformer, which may act either inductively or capacitatively, there is a risk that for a given frequency the negative feedback aimed at becomes a positive feedback. To prevent instability, the feedback thus cannot be made as strong as might be desired.

As a consequence of all the difficulties mentioned above one is compelled to use a heavy and large (and hence expensive) output transformer, unless lower demands are made.

Some years ago Philips started working in another direction, viz. avoiding an output transformer altogether. If this attempt were to succeed, it would be possible, not only to eliminate all these drawbacks, but also to get rid of an expensive, bulky and heavy component. That these attempts have, in fact, been successful will appear from the following.

High resistance loudspeakers

Loudspeaker speech-coils having few turns of fairly thick wire have the advantage of being quickly wound, with little risk of breaking the wire, and also that the insulation of the wire takes up relatively little winding space (high filling factor).

The first investigations were aimed at establishing the largest number of turns and the smallest wire diameter that could be accommodated in the available space without unduly complicating manufacture or making the filling factor too small. Supported by their experience with all kinds of miniature

[3]) See the article quoted in [1]), page 314.

winding jobs, the Eindhoven loudspeaker factory succeeded in 1953 in mass-producing 4000 Ω coils using 40 μ copper wire. These coils were provided with a central tapping, so that they could be directly incorporated in a push-pull circuit (*fig. 1*). The

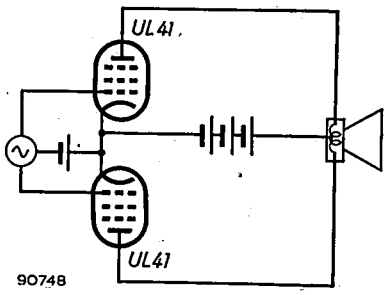


Fig. 1. Output stage with two pentodes UL 41 and a split-coil loudspeaker (type 9754) without output transformer. The resistance of the speech coil is 2×2000 ohms.

resistance of 2000 Ω for each half coil is fairly low for ordinary output valves, but is a practicable value for the UL 41 valve with its rather low D.C. resistance [4]). Both direct tests and life tests proved satisfactory, and part of the production of the radio receiver BX 521 U was equipped with such a loudspeaker, so that practical experience on a large scale could be gained. As regards the loudspeakers, no complaints were received. A remarkable feature of this set was the improved reproduction of the low notes.

This solution, however, has its drawback. The output valves are operated as a Class A amplifier, i.e. with a large anode direct current. This current flows through the high-resistance loudspeaker coil, causing a considerable voltage drop (far greater than in an output transformer), and a considerable development of heat. In a new circuit, described below, these disadvantages have been completely eliminated.

The single-ended push-pull circuit

Fig. 2a shows a normal push-pull circuit. Each of the valves has been biassed to give a quiescent anode current $I_{a(w)}$. The supply, at a voltage V_b, therefore delivers a power $2V_bI_{a(w)}$. Fig. 2*b* shows a somewhat modified circuit, in which each valve again has the supply voltage V_b and the current $I_{a(w)}$, the total D.C. power being likewise $2V_bI_{a(w)}$. If the load resistance R_0 is the same in either case, the efficiency is also the same and the circuit of fig. 2*b* can supply the same A.C. power as that of fig. 2*a*.

This remains valid if PP' and QQ' in fig. 2*b* are interconnected. The two resistances R_0 are then connected in parallel and are equivalent to a single resistance $\frac{1}{2}R_0$ (fig. 2*c*), i.e. $\frac{1}{4}$ of the total resistance $2R_0$ of the circuit in fig. 2*a*. If the two valves are equally biassed, direct current flows only through the circuit formed by the supply source (voltage $2V_b$) and the two series-connected valves, none passing through the resistance $\frac{1}{2}R_0$.

This type of push-pull circuit was already well known in 1951 [5]). Clearly, it is ideally suited to the system of direct power transfer: it requires a 4 times lower load impedance than the conventional push-pull circuit, whilst owing to the fact that the load is free of direct current, all drawbacks such as large

[4]) Valves of the U-series are designed for universal A.C./D.C. receivers without mains transformer. They therefore have a low anode voltage (100-170 V), so that output valves of the U-series have to be designed for handling a large current. Their D.C. resistance R_{ak} is accordingly low.

[5]) A. Peterson and D. B. Sinclair, A single-ended push-pull audio amplifier, Proc. Inst. Rad. Engrs. **40**, 7-11, Jan. 1952.

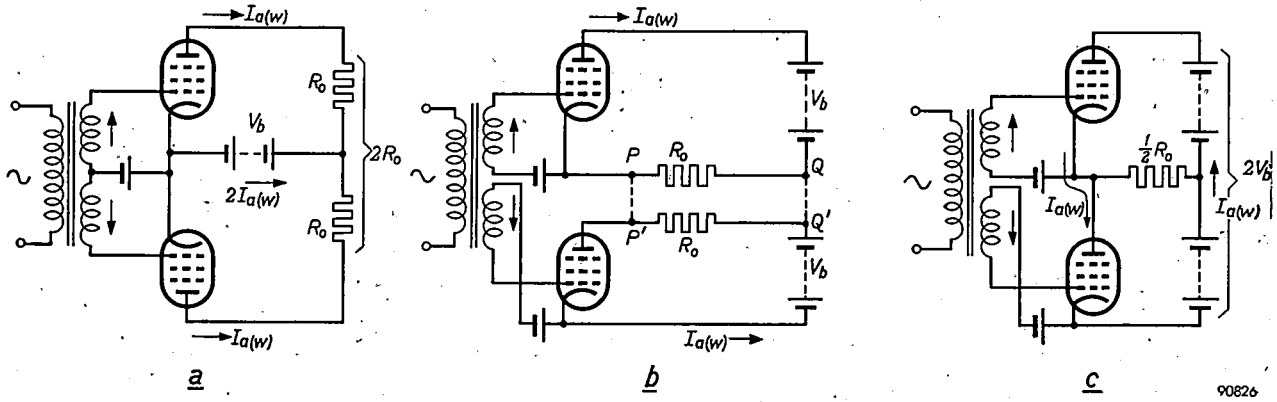


Fig. 2. *a*) Conventional push-pull circuit. *b*) Circuit which is electrically equivalent to (*a*). By interconnecting P and P' and Q and Q', the single-ended push-pull circuit (*c*) is obtained, which matches to a 4 times lower load resistance ($\frac{1}{2}R_0$ instead of $2R_0$), whilst no D.C. flows through the load.

voltage losses and heat development in the loudspeaker coil are eliminated. The latter, moreover, does not require a central tapping, which simplifies manufacture. The central tapping of the supply source can likewise be eliminated, e.g. by connecting the loudspeaker L in series with a capacitor C_0 of sufficiently large capacitance according to *fig. 3*. Another method will be discussed presently.

The above principle has been applied in an increasing number of Philips receivers during the last few years. A new output pentode (type EL 86) has been developed for it, which can operate at low anode voltage and large anode current [6]); at maximum anode dissipation, the D.C. resistance R_{ak} amounts to only 1600 Ω. For the two valves, this requires matching by a loudspeaker impedance (mainly resistive) of 800 Ω, which presents no difficulties in manufacture. A single-ended push-pull

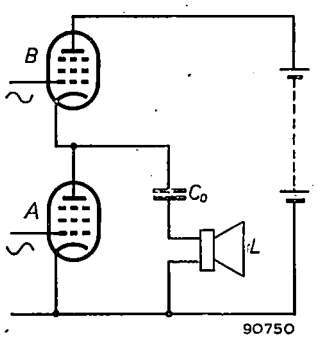


Fig. 3. Single-ended push-pull circuit with two pentodes A, B (EL 86) requiring no central tapping of the H.T. supply. A large capacitor C_0 is placed in series with the speech coil of loudspeaker L.

circuit with two EL 86 valves can produce a maximum A.C. power of 10 W; for 10-20 W two EL 86 valves may be connected in parallel on either side, the required load resistance then being 400 Ω, which may be realized by using one 400 Ω loudspeaker or two 800 Ω loudspeakers in parallel. Nearly all Philips loudspeakers with a cone diameter of 12.5 cm and larger are now available in 400 or 800 Ω versions [7]).

The relations between load resistance R_0, output power P_0 and the position of the operating point $V_{a(w)}$, $I_{a(w)}$ can be determined from the following simplified considerations. *Fig. 4* again shows the circuit of fig. 3 but with voltage and current indications added. The load resistance is here denoted R_0; the reactance of C_0 will be neglected. *Fig. 5* shows two idealized I_a-V_a characteristics of the pentodes employed. One characteristic is valid for a control-grid voltage $V_{g1} = 0$ and the other for the value of V_{g1} where $I_{a(w)}$ is half as large. XY is the load line intersecting the knee of the upper characteristic and subtending an angle α with the V_a axis such that $\cot\alpha = 2R_0$, this being the most favourable load line (i.e. a load of $2R_0$ for each valve; being in parallel, this gives a net load of

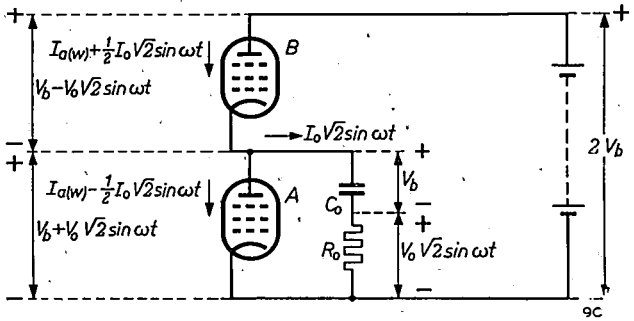


Fig. 4. Circuit of fig. 3 showing currents and voltages. A load resistance R_0 represents the loudspeaker.

R_0 as in fig. 4). It intersects the V_a axis at the point Y, corresponding to the voltage $2V_b$ of the supply source. The working point is W. We see that the relation $\cot\alpha = (2V_b - V_{a(w)})/I_{a(w)}$ is also valid; by neglecting the "knee voltage" $V_{a\,min}$ below which V_a must not drop, we have $2V_b = 2V_{a(w)}$ and therefore $\cot\alpha = V_{a(w)}/I_{a(w)}$. By definition, $\cot\alpha = 2R_0$ and $V_{a(w)}/I_{a(w)} = R_{ak}$, so that the matching condition $2R_0 = R_{ak}$ is satisfied.

The r.m.s. values of the alternating current I_0 and the alternating voltage V_0 are given by:

$$I_0 = \sqrt{P_0/R_0}, \quad V_0 = \sqrt{P_0 R_0}. \quad \ldots\ldots \quad (1)$$

Either valve contributes half the amount of I_0. The valve currents are $I_{a(w)} \pm \frac{1}{2}I_0\sqrt{2}\sin\omega t$ (see fig. 4). Under conditions of full drive (maximum signal), the entire load line XY is traversed, I_a swinging from zero to $2I_{a(w)}$. Therefore

$$I_{a(w)} = \tfrac{1}{2}I_0\sqrt{2}. \quad \ldots\ldots\ldots \quad (2)$$

V_a swings from $V_{a\,min}$ (approx. zero) to $2V_b$, so that

$$V_{a(w)} = V_0\sqrt{2}. \quad \ldots\ldots\ldots \quad (3)$$

For the voltage $2V_b$ of the H.T. supply, we thus have

$$2V_b = 2V_{a(w)} = 2V_0\sqrt{2}. \quad \ldots\ldots \quad (4)$$

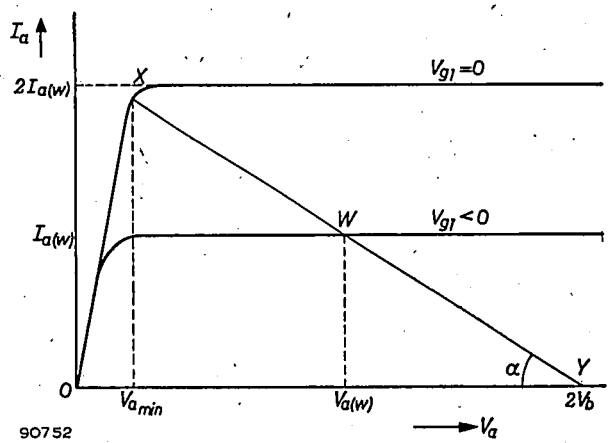


Fig. 5. Idealized pentode characteristics and load line XY. The working point is W.

[6]) As was recently pointed out in this Review (K. Rodenhuis *et al.*, Philips tech. Rev. **18**, 185, 1956/57, No. 7), such conditions are more favourable to the working life of the valve than when the same dissipation is obtained at a higher voltage and a smaller current.

[7]) 800 Ω speakers are designated by the suffix A, 400 Ω speakers by the suffix B to the type number, e.g. 9710 A, 9766 BM; the suffix M indicating a double-cone loudspeaker (see J. J. Schurink, Philips tech. Rev. **16**, 241-249, 1954/55, or the article quoted by [1]), pp. 305-306).

The power P_b actually delivered by the source is

$$P_b = 2V_b I_{a(w)} = 2V_0\sqrt{2} \times \tfrac{1}{2}I_0\sqrt{2} = 2V_0 I_0 = 2P_0,$$

which demonstrates that the efficiency is 50%.

In the *absence* of a signal the supply delivers the same power $P_b = 2P_0$, which is now entirely dissipated in the valves, which means for each valve a dissipated power $P_0 = V_b I_{a(w)}$ (twice as much as under conditions of full drive). This means that the maximum A.C. power the circuit is permitted to deliver, is equal to the maximum anode dissipation in the absence of a signal. This quantity is a parameter of the valve, and amounts to 12 W for the EL 86. Owing to the simplifications introduced here, P_0 is actually somewhat lower, viz. 10 W.

For $P_0 = 10$ W and $R_0 = 800\ \Omega$, we arrive, according to (1), at $I_0 = 112$ mA and $V_0 = 90$ V. The working point of the valve now follows from (2) and (3):

$$I_{a(w)} = 79 \text{ mA}, \quad V_{a(w)} = 127 \text{ V}.$$

According to (4), the supply voltage must amount to $2V_b = 254$ V, to which must be added the amount $2V_{a\,min}$ (and, if the negative grid bias of valve A in fig. 4 is produced across a cathode resistor, an additional $|V_{g1}|$).

For maximum output powers lower than 10 W the valve may be biassed to give a smaller quiescent anode current at a proportionally lower supply voltage, the working point being evaluated in the above manner.

The signal to be amplified generally consists of a single alternating voltage with respect to earth. For driving the valves of a push-pull circuit in opposite phase, however, two control voltages are required. This could be obtained by means of a transformer with two secondary windings as shown in fig. 2*a* for the ordinary push-pull circuit and in fig. 2*c* for the single-ended push-pull circuit. For some time however, a valve (phase inverter) has been preferred to obtain the opposed control voltages for a conventional push-pull circuit. A valve can also be used for this purpose with a single-ended push-pull circuit, as is shown in *fig. 6*. Section I of a double-triode ECC 83 functions here as amplifier and is connected to the control grid of the output valve A (for the sake of clarity, coupling capacitors are not shown in the diagram) and also to the grid of section II. At the anode of II a signal in opposite phase is produced, which, by an appropriate choice of resistances, is given an amplitude suitable for driving output valve B.

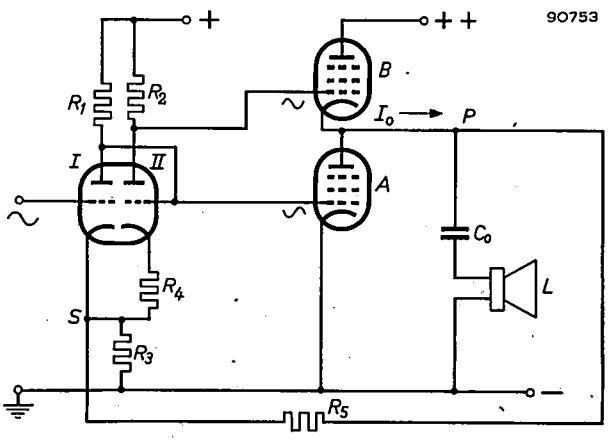


Fig. 6. *I-II* double-triode ECC 83. *A*, *B* pentodes EL 86 in single-ended push-pull arrangement. The first stage of the amplifier, consisting of triode *I*, drives the output valve *A* and the phase inverter *II* producing the control voltage for *B*. (Components not essential to the principle, such as coupling capacitors, are not shown.) Positive feedback in the first-amplifier - phase-inverter stage is effected via the common cathode resistor R_3 of *I* and *II*. Negative feedback, from *P* to *S*, is obtained via R_5.

Feedback

As mentioned earlier, one of the advantages of omitting an output transformer is the greater freedom in applying negative feedback. Without running the risk of instability, certain combinations of negative and positive feedback can even be made, by which, as we shall demonstrate presently, distortion can be drastically reduced. In the circuit of fig. 6 [8]) negative feedback is realized by passing part of the output current I_0 (being proportional to the output voltage), via resistor R_5, through cathode resistor R_3; at the same time there is positive feedback in the phase inverter stage in the form of valves I and II having resistor R_3 as a common cathode resistance.

The effect of combined positive and negative feedback is clearly demonstrated by a hypothetical example of greater simplicity than that of fig. 6. A_1 and A_2 in *fig. 7* represent two amplifiers in cascade, e.g. a pre-amplifying stage and an output stage, whose amplifications will likewise be called A_1 and A_2 respectively. A portion B_1V_1 of the output signal V_1 of A_1 is fed back to the input of A_1, and a portion B_2V_0 of the output signal V_0 of A_2 is likewise fed back to the input of A_1. If V_i is the signal to be amplified, we may write, quite generally,

$$V_1 = A_1(V_i + B_1V_1 + B_2V_0)$$

and

$$V_0 = A_2V_1.$$

The overall amplification A is accordingly:

$$A = \frac{V_0}{V_i} = \frac{A_1A_2}{1 - A_1B_1 - A_1A_2B_2} = \frac{A_1A_2}{N},$$

where

$$N = 1 - A_1B_1 - A_1A_2B_2.$$

Similarly, the total distortion D is given by

$$D = \frac{1}{N}D_1 + \frac{1 - A_1B_1}{N}D_2 + \frac{1 - A_1B_1}{N}D_1D_2,$$

[8]) Designed by E. H. Nielsen, formerly a member of the laboratory of the Philips factory in Copenhagen.

where D_1 and D_2 are respectively the distortion of either amplifier.

It may be seen that if A_1B_1 is made unity, a very special situation arises: D is reduced to D_1/N, i.e. the distortion of amplifier A_2 does not contribute

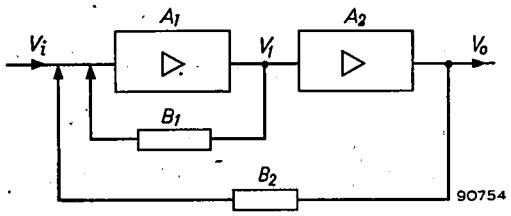


Fig. 7. Cascade connection of two amplifiers A_1 and A_2, the former having a feedback loop via B_1, and the combination a feedback via B_2.

at all to the total distortion. The remaining distortion, D_1/N, is determined by the (slight) distortion D_1 of the first amplifier and by the quantity N, the latter becoming $-A_1A_2B_2$ for $A_1B_1 = 1$. The absolute value of $A_1A_2B_2$ may be considerable larger than unity, in which case the total distortion D will even be appreciably smaller than D_1. Also, the absolute value of the total amplification A becomes equal to $1/B_2$ (and therefore independent of A_2).

The condition $A_1B_1 = 1$ can be fulfilled for a wide frequency range if A_1 and B_1 are real in that range, i.e. if the amplifier A_1 and the feedback circuit B_1 contain no phase-shifting elements. In the first amplifying stage this condition may be closely approached without any difficulty. This condition being fulfilled means that the amplifier A_1 is given such a positive feedback B_1 that it is on the verge of oscillating. This does not necessarily mean that the combination A_1-B_1-A_2-B_2 (fig. 7) is unstable; if the second amplifier and the feedback circuit are free of any elements causing adverse phase shifts, the whole circuit can be kept stable by applying a certain negative feedback B_2. An output transformer is an element which inevitably causes such phase shifts that the stability would be seriously jeopardized [9]). Only by its elimination is it possible to realize circuits of the type considered here.

Practical examples

A.F. amplifier for a radio receiver

Fig. 8 shows a further developed version of the circuit of fig. 6, as it might be used for a radio receiver. Here the feedback resistor R_5 of fig. 6 has been replaced by two networks, C_1-R_6-R_7 and C_2-R_8-R_9. These incorporate the tone controls: R_6 for the low and R_9 for the high notes.

[9]) The output transformer can be kept outside the feedback loop B_2 (see e.g. fig. 6 of the article quoted by [5])), but the distortion introduced by it remains undiminished in the output signal; also the degree of negative feedback is necessarily limited owing to the phase shift between primary voltage and primary current.

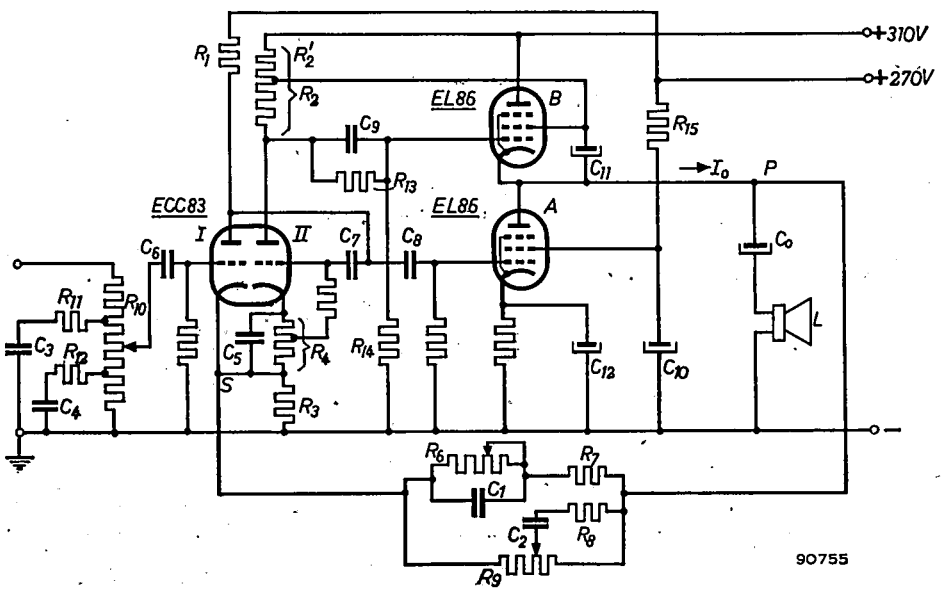


Fig. 8. More detailed version of the schematic diagram of fig. 6. Instead of the feedback resistor R_5, the networks C_1-R_6-R_7 and C_2-R_8-R_9 have been incorporated; R_6 low-note control, R_9 high-note control. R_{10} volume control with branches R_{11}-C_3 and R_{12}-C_4 for low-note compensation at low levels of volume. C_5 reduces the negative feedback through R_4 for the high audio frequencies. C_6, C_7, C_8, C_9 coupling capacitors. The voltage divider R_2-R_{13}-R_{14} provides the correct bias for the control grid of output valve B. The screen grid of the output valve A is fed via resistor R_{15} and the screen grid of B from a tapping on R_2 (for other methods of screen-grid supply, see figs. 11 and 12).

Low notes. Capacitor C_1 reduces the negative feedback at low frequencies, so that the latter are relatively more amplified. For a radio set this is a desirable feature, e.g. for compensating the drop in loudspeaker radiation at frequencies below those at which the cabinet forms a sufficiently large baffle. With a cabinet of average size, the drop in loudspeaker response at descending frequency (6 dB per octave) begins at about 1000 c/s. As desired, the effect of C_1 can be decreased (increasing the negative feedback and thus increasing low-note attenuation) by lowering the resistance of R_6. The non-variable resistor R_7 prevents the negative feedback becoming too strong.

has been compensated by placing a capacitor C_5 across cathode resistor R_4 of the phase inverter. The effect of this is a reduction of the negative feedback in the phase-inverter stage, thus increasing the gain, particularly at the higher frequencies.

Figs. 9 and *10* give some idea of the results attained with an amplifier of the above type. The diagrams of fig. 9 show the distortion as a function of the output power for the frequencies 90, 1000 and 8000 c/s, whilst fig. 10 shows the amplitude-frequency characteristic with both tone controls set to maximum. It can been seen that the low and high notes are boosted with respect to the medium range, as is desirable for a radio receiver.

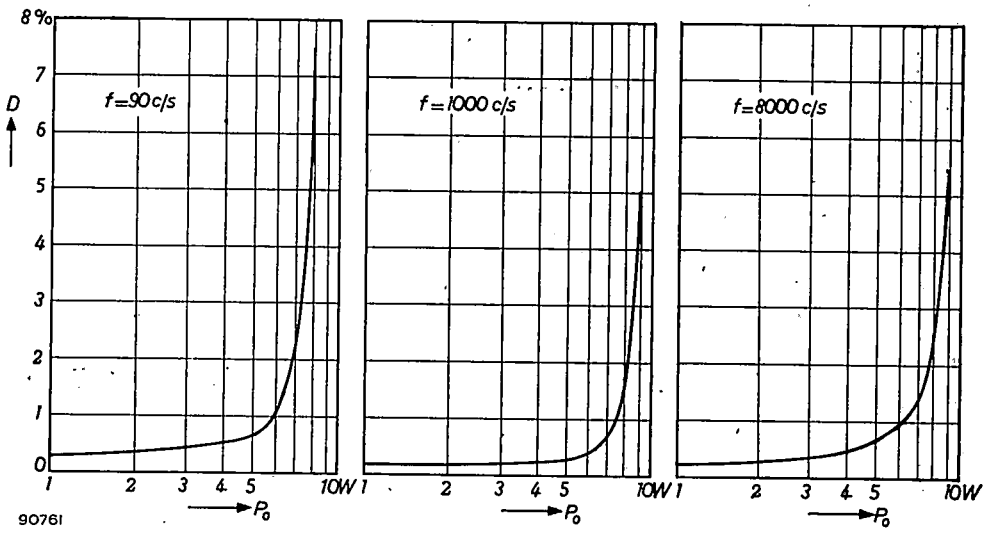


Fig. 9. Distortion D, measured on an amplifier of the type shown in fig. 8, plotted against the output power P_0, at the frequencies 90, 1000 and 8000 c/s. Both tone controls set to maximum.

High notes. The values of C_2 and R_8 are so chosen that the higher audio frequencies are mainly fed back through this branch. If the sliding contact on R_9 in fig. 8 is moved to the right, a larger part of R_9 is connected in series with this branch, the negative feedback diminishes and the high notes become accordingly stronger.

As is customary in many Philips receivers, automatic low-note compensation is provided at low levels of volume: volume control R_{10} is provided with the branches R_{11}-C_3 and R_{12}-C_4, which ensure that with a reduction of the total volume the low notes are relatively less attenuated, thus compensating for the property of the human ear of becoming insensitive to a steadily increasing portion of the low-note range when the volume is reduced.

Near the upper limit of the audio spectrum the amplification is reduced as a result of the anode capacitance of the output valves and of the Miller effect (reaction of anode upon control grid). This

As a measure of the sensitivity of an amplifier, it is customary to give the input r.m.s. voltage required to produce an output power of 50 mW at 1000 c/s. For this particular example this voltage amounts to 24 mV, which is normal for the audio section of a radio receiver.

Negative feedback reduces the internal resistance of the output stage to about 20 Ω, a value that is

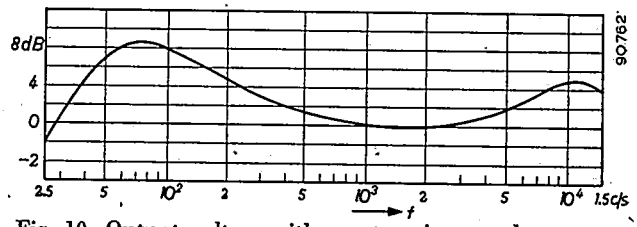


Fig. 10. Output voltage with constant input voltage versus frequency f, measured on an amplifier of the type shown in fig. 8, loaded with 800 ohms. Both tone controls set to maximum (i.e. minimum negative feedback). Boosting of the low and high notes as shown is desirable in the audio amplifier of a radio receiver.

low compared with the load resistance (800 Ω). This is conducive to the uniform reproduction of the very low notes (damping of the loudspeaker resonance).

With a single-ended push-pull circuit a question requiring attention is the supply for the screen grids of the output valves. The screen grid voltage should be at about the same D.C. level as the anode, but there must be no alternating voltage between screen grid and cathode.

Let us first investigate how this ideal condition can be approached with the output valve A (fig. 8). Here the screen grid is supplied via resistor R_{15}, whilst the electrolytic capacitor C_{10} (in series with C_{12}) constitutes a low impedance between screen grid and cathode. The screen-grid current (about 5 mA) forms only a minor portion of the average cathode current. This portion is not constant however: it varies with the amplitude of the signal on the control grid. The average value of the screen-grid current therefore varies with the signal strength, so that the presence of the series resistor R_{15} makes the screen-grid voltage vary with signal strength: the stronger the signal, the lower the screen-grid voltage. This effect reduces the maximum output power of the valve. To minimize the effect, the resistance R_{15} should be made as small as possible and the screen grid must therefore be fed from a source whose voltage exceeds the required value by as little as possible. For this reason the screen grid in fig. 8 is not fed from the 310 V source, but from the 270 V source (which also supplies the valve-half I and the R.F. part of the receiver). A supply direct from a voltage of approx. 155 V would be even better, and this can be easily realized if the 310 V supply is derived from a full-wave bridge circuit rectifier (*fig. 11*): if the secondary of the transformer is provided with a central tapping, half the direct voltage can be tapped off from this point.

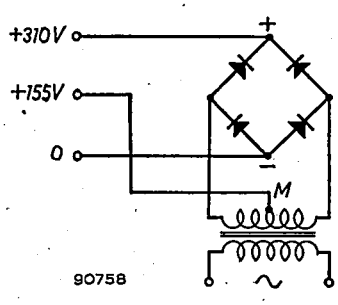


Fig. 11. Full-wave bridge circuit rectifier. Half the direct voltage can be obtained from the centre-tapping M on the secondary of the transformer. This can be used for feeding the screen grid of valve A in fig. 8.

As for the output valve B (fig. 8), its anode has a fixed potential, whilst the output signal voltage is taken off at the cathode. If the screen grid were connected directly to the anode, then the loudspeaker would be nearly short-circuited (for signal currents) via capacitor C_{11} and the H.T. supply. A certain impedance between screen grid and anode must therefore always be provided (unless a separate power supply is to be provided for this screen grid). This impedance should be high with respect to the (parallel) loudspeaker impedance, but, on the other hand, it should not have a high D.C. resistance, as otherwise the screen-grid voltage would be too much below the anode voltage and also would vary too much with the signal strength (see above). In some respects a choke has some advantages over a resistor, but this would again introduce a phase-shifting element. In fig. 8 a compromise has been struck between a not too high and not too low resistance by connecting the screen grid to a suitable tapping on the anode resistor R_2 of the phase inverter.

A more elegant solution is shown in *fig. 12*. Here the loudspeaker itself is connected between screen grid and anode. The output signal current I_0 passes

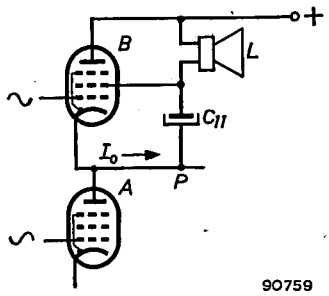


Fig. 12. Supply to the screen grid of valve B in fig. 8 via the loudspeaker L. The output current I_0 (less the negative-feedback current) flows through C_{11} and L.

without losses (with the exception of the part used for feedback) via C_{11} through the loudspeaker. The function of capacitor C_0 is now taken over by C_{11} and it can therefore be omitted. The fact that the (small) screen-grid current now flows through the loudspeaker forms no objection.

Further improved circuit

The reader may have noted that in fig. 8, from which the output transformer with its phase-shifting effect has been eliminated, other phase-shifting elements have after all been included, viz. the tone control networks in the feedback circuit. In principle such a solution is indeed not ideal. That it is nevertheless used is justified by the fact that otherwise an additional amplifying stage would be required

(the additional amplification of low and high notes is effected here by reducing the negative feedback), and by the fact that even with the smallest amount of negative feedback, distortion remains appreciably smaller than in a conventional circuit with the average output transformer.

If near-perfection is aimed at, then the negative feedback loop should be kept free from any phase shift by effecting the tone control in an additional stage preceding valve *I*. For an amplifier of this type the following distortion figures were measured:

Output power (in W)	10.4	9.25	4	1
Distortion (in %)	0.30	0.11	0.05	0.03

In order to obtain for this measurement an input signal itself sufficiently free of distortion, a filter for suppressing harmonics should be incorporated between the audio generator and the amplifier. Any harmonics in the output signal of the amplifier are measured with a "wave analyzer", by means of which each harmonic can be separately determined.

Fig. 13 shows the amplitude and phase characteristics of this amplifier. We see that the point where the amplification has dropped by 3 dB lies well into the ultrasonic range, namely at 250 kc/s, whilst the phase shift in the audible range is restricted to a few degrees only. For an output power of 10 W the efficiency of the output stage is a good 42%, which is very satisfactory. The only critical resistance value is that of the cathode resistor of the phase inverter (R_4), which must be accurate within 2%; for all other resistors a 10% tolerance is adequate.

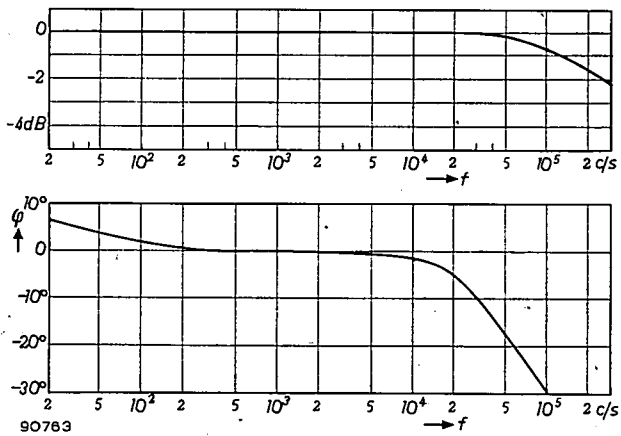


Fig. 13. Amplitude characteristic (above) and phase characteristic (below) of an improved single-ended push-pull amplifier.

We shall confine ourselves to the above two examples. It may be mentioned, however, that in many other cases the greater freedom in the application of negative feedback which single-ended push-pull makes possible will be welcome: it permits the realization of various types of special circuits for which the output transformer has hitherto formed the great stumbling block.

Summary. The low resistance of the speech coils of conventional moving-coil loudspeakers requires a transformer for matching to the far higher D.C. resistance of the output pentode. This output transformer has many disadvantages: it limits the reproduced audio spectrum on either end, causes power losses, introduces distortion, and by the phase-shift it introduces, prevents negative feedback being employed to its full advantage. Any transformer in which these drawbacks are confined within acceptable limits is necessarily a heavy, bulky and expensive component.

In 1953 Philips manufactured loudspeakers with a speech coil of 2× 2000 ohms, which could be incorporated in a normal push-pull circuit with two pentodes UL 41. For such a "direct power transfer", however, the "single-ended push-pull circuit" is more suitable: here the load (loudspeaker) can have a resistance 4 times as low as with the conventional push-pull circuit, whilst no direct current flows through it. This led to the development of the output pentode EL 86, with a D.C. resistance of 1600 ohms at full load, and of associated loudspeakers of 800 and 400 ohms. Two valves EL 86 in single-ended push-pull can deliver 10 W.

The absence of an output transformer makes it possible to realize circuits in which the first amplifying stage has positive feedback up to the verge of oscillation, whilst the first amplifying stage and the output stage together have negative feedback. Such a circuit is perfectly stable. The total distortion can be reduced to only a small fraction of the distortion of the first amplifying stage alone and made independent of the distortion in the output stage. Two practical examples are discussed, one with the tone controls incorporated in the negative feedback circuit and one with the tone controls in a separate preceding stage. In the latter case the distortion at 10 W is only 0.30%; the amplitude characteristic remains straight well into the ultrasonic range before showing a 3 dB drop at 250 kc/s, so that the phase characteristic does not show deviations greater than a few degrees within the audible range.

STROBOSCOPIC OPERATION OF PHOTOMULTIPLIER TUBES

by C. F. HENDEE *) and W. B. BROWN *).

621.383.27

By applying pulse voltages instead of the usual D.C. voltages to the electrodes of a photomultiplier tube, such a tube is made to operate as a light shutter capable of shutter times of fractions of a microsecond. An interesting method for the analysis of very weak light flashes and for "time-resolved spectroscopy" is based on this type of operation. The behavior of the pulsed photomultiplier tube is in several respects different from that under D.C. operation.

Aim and principles of method

Multiplier phototubes, in which the electron current emitted by a photocathode under illumination is enormously amplified by secondary electron emission from a number of successive electrodes ("dynodes"), have found wide-spread application in modern physics and technology. A specialized field of great importance is scintillation counting, in which high energy particles or radiation quanta are detected, counted and their energy measured by means of the flash of light they produce in a scintillating crystal or liquid [1]). Equally important is the direct measurement of weak light sources, such as stars, spectral lines, etc. and of rapidly varying light sources, for instance in sound and facsimile reproduction [2]). The main characteristics utilized in such applications are the extremely high sensitivity of the photomultiplier tube and its extremely quick response: the time delay, given by the total transit time of electrons through the tube, amounts to a few 10^{-8} sec.

A special problem that will be discussed in this article and to which the photomultiplier tube is ideally suited, is the spectral study of weak light flashes, e.g. from electrical discharges in gases. Generally this has been done in the past by presenting the amplified output of a photomultiplier tube on an oscilloscope synchronized with the light flash. The oscilloscope trace will then indicate the time variation of the light intensity, either of the total light or of that in a specified spectral region if this is singled out by a monochromator and focused on to the photomultiplier tube.

With weak sources, and especially after spectral dispersion that may reduce the intensity to a minute fraction, one ultimately reaches in these studies a point where the whole signal to be detected and displayed is carried by a few photoelectrons [3]). This is clearly seen in *fig. 1* which shows that for the lowest light intensities the gradually varying magnitude of the signal is not represented by a gradually varying vertical displacement of the oscilloscope trace but rather by a variation of the "density" of separate photoelectron pips. The determination of the relative intensities and pulse shapes from such signals obviously will be rather difficult and the accuracy very poor.

The photomultiplier tube can be made to provide a very effective solution for this problem in the following manner. Since the tube is sensitive to light only when proper voltages are applied to the cathode and the dynodes, it can be made to operate as a *light-shutter* by applying these voltages during a well-defined period, i.e. in the form of a rectangular pulse. The "on" time of the tube can be restricted to say 10^{-7} sec and can be timed by suitable pulse techniques to occur at any desired phase of the light flash. The photocurrent obtained at the collector of the tube during the "on" time is fed to a D.C. microammeter, preferably of a recording type. If now the light flash is periodically repeated and the photomultiplier tube is pulsed always in the same relative phase, the pulsed photocurrent may be *integrated* over a suitably long period by having a long time constant in the current meter, without loss of information concerning the time variation of the light flash.

It will be recognized that this procedure is similar to the obtaining of a stationary image of periodic

*) Philips Laboratories, Irvington-on-Hudson, N.Y., U.S.A.

1) G. A. Morton, Photomultipliers for scintillation counting, RCA Rev. **10**, 525-553, 1949.
A. Krebs, Szintillationszähler, Ergebn. exakt. Naturwiss. **27**, 361-409, 1953.

2) See for instance R. W. Engstrom, Multiplier photo-tube characteristics: application to low light levels, J. Opt. Soc. Amer. **37**, 420-431, 1947.

3) R. F. Saxe, Detection and recording of optical transients by photo-multipliers, Nature (London) **172**, 1198, 1953.

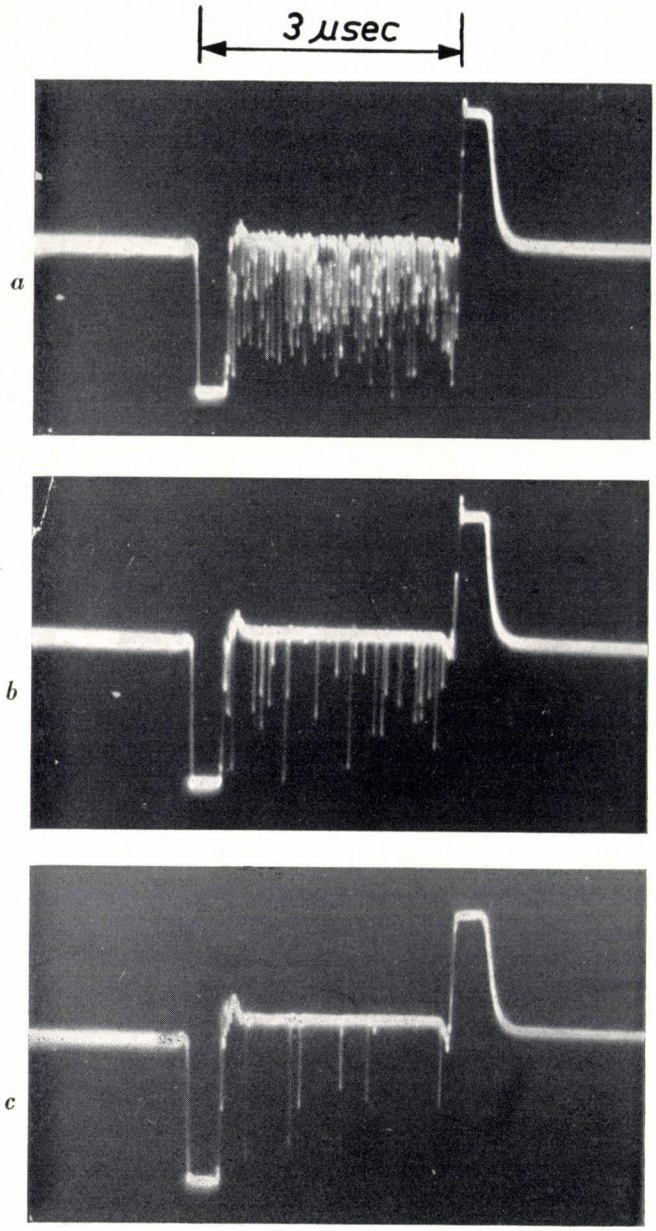


Fig. 1. Oscilloscope traces of the output of a photomultiplier tube subject to a very weak illumination, decreasing in intensity from (*a*) to (*c*). For reasons discussed later, the photomultiplier tube in each case was made to operate only during a short time, viz. 3 microseconds. During this time, only a few photoelectrons are produced with the lowest intensities. Owing to the random nature of the photon arrival and of the secondary electron emission process in the tube, the pips caused by separate photoelectrons occur at irregular intervals and differ greatly in amplitude.

The two large pips (one down on the left and the other up on the right of each trace) in this figure and also in figs. 8 and 9 result from capacitive pick-up at the photo-tube collector of the high-voltage pulse applied to the dynodes (cf. fig. 10).

processes by stroboscopic observation [4]). The analogy is made complete when the phase of the photomultiplier tube is gradually shifted so as to *scan* the complete light pulse.

Apart from the ease of evaluation, the method provides a considerable increase in accuracy as compared with the measurement of a single flash: for very low intensity light, where the primary photocurrent in each pulse may consist of only a few (or even less than one, on the average) photoelectrons, the statistical variation of this number of photoelectrons and hence the probable error in the determination of intensity is relative large. Using repetition rates of the light flash of up to 10 000 per second and integration times up to 10 seconds, a number of photoelectrons 10^5 times larger than in a single flash can be accumulated, and the accuracy thereby improved by a factor of about 300. A more detailed discussion of this improvement will be given below.

Evidently the method can be used only for light sources which can give repetitive flashes. Many types of electrical discharges in gases permit periodic pulsing at high frequencies [5]). Among the subjects investigated by this technique in the Philips Laboratories at Irvington is the light from a hydrogen thyratron and the light from low-pressure discharges in nitrogen. In most cases light flashes of 1 microsecond duration were used. The set-up for the measurements, using a monochromator, is shown schematically in *fig. 2*; a photograph of the equip-

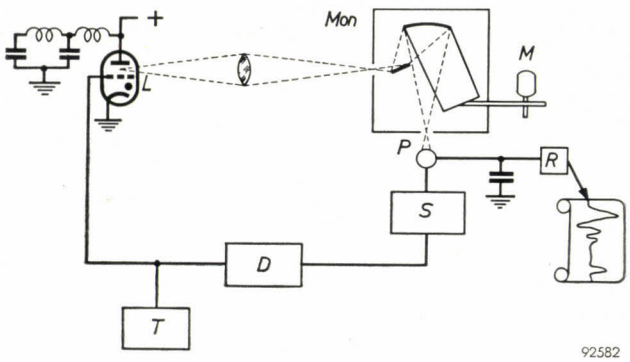


Fig. 2. Set-up for time-resolved spectroscopy using a stroboscopically operated photomultiplier tube. *T* trigger circuit, *L* pulsed light source to be analyzed (in this case a hydrogen thyratron), *Mon* grating monochromator with motor drive *M*, *S* circuit providing the pulse voltages for the photomultiplier tube *P*, *D* variable delay circuit, *R* recording D.C. microammeter.

In most experiments described in this article the RCA photomultiplier tube 1 P 21 was used.

---

4) C. F. Hendee, Time-resolved spectroscopy with pulsed photomultipliers, J. Opt. Soc. Amer. **43**, 330, 1953 (abstract). C. F. Hendee and W. B. Brown, Time-resolved spectroscopy of nitrogen and hydrogen, Phys. Rev. **93**, 651, 1954 (abstract).

A similar technique for longer pulses is described by A.V. Phelps and J. L. Pack, Measurement of time varying optical absorption, Rev. sci. Instr. **26**, 45-49, 1955.

5) C. F. Hendee, Band spectra in pulsed nitrogen discharges, Phys. Rev. **84**, 1075, 1951 (abstract). Cf. also R. D. Drosd, Gated multiple light pulse generator, Bull. Amer. Phys. Soc. **30**, No. 4, p. 36, 1955 (abstract).

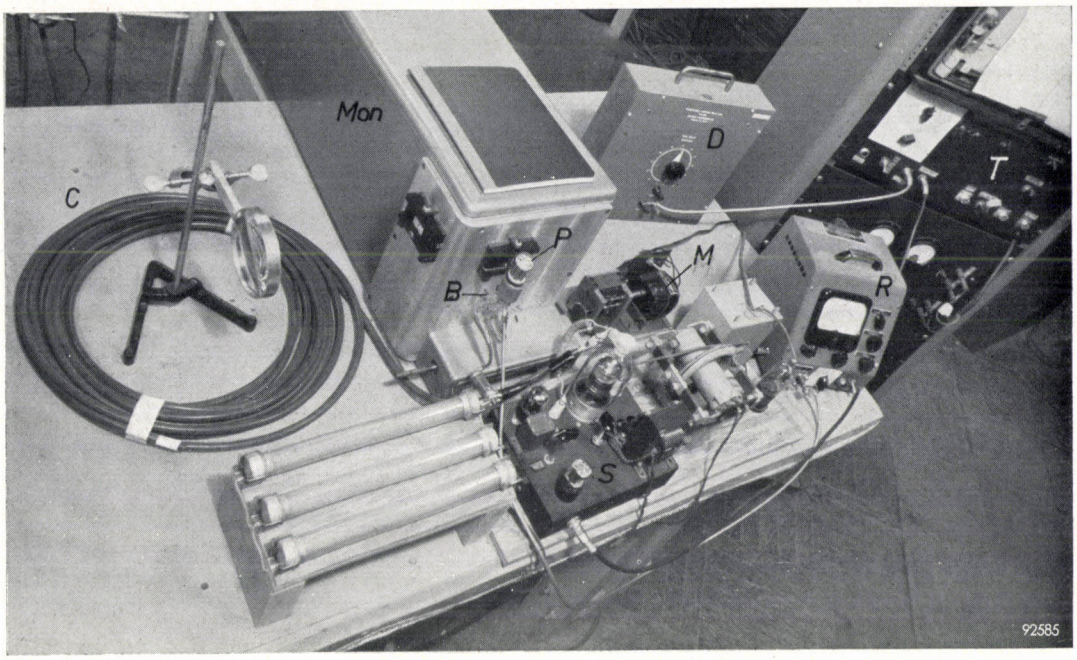
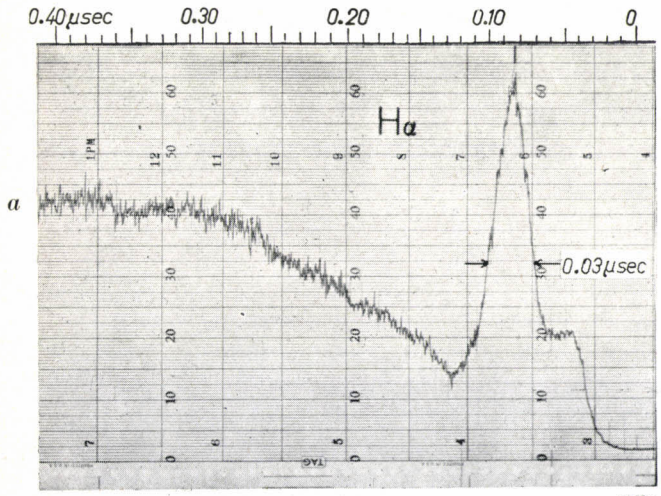
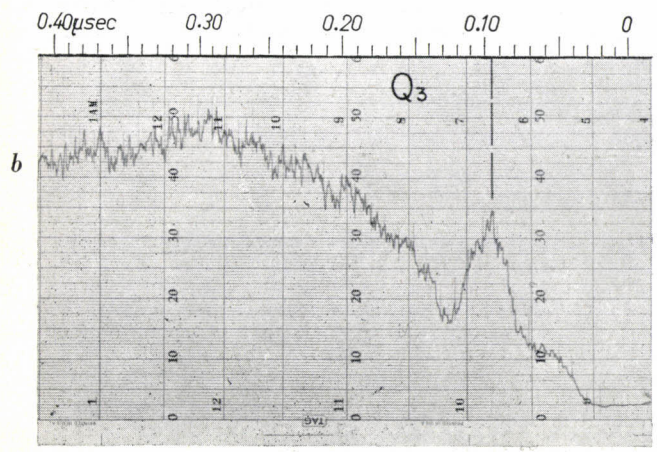


Fig. 3. Photograph of equipment outlined in fig. 2. Same lettering as in fig. 2. The light source is to the left (not visible). To the right, a rack containing the trigger circuit T and the recording part of the current meter R is seen. C is a length of coaxial cable acting as a pulse forming network in the circuit S. B are the voltage-divider resistors for the photomultiplier-tube dynodes (see fig. 10).



a



b

ment is seen in *fig. 3*. In order to determine the time variation of the intensity of a particular spectral line during a flash, the monochromator setting was fixed on this line and the short photomultiplier pulse was slowly phased through the light flash by means of the variable delay circuit, D in fig. 2. A record is thus obtained on the strip chart recorder which represents the light intensity as a function of time [6]). *Fig. 4* is such a record showing part of the pulse shape of two hydrogen lines in the light emitted by a pulsed hydrogen thyratron. Alternatively a complete spectrum of the light can be recorded at particular instants of the flash time with an approp-

[6]) Proper attention, of course, must be given to the selection of the scanning time; the problem is similar to that described in some detail in this Review for the analogous case of the scanning of X-ray diffraction patterns: see W. Parrish, X-ray intensity measurements with counter tubes, Philips tech. Rev. **17**, 206-221, 1955/56.

Fig. 4. Part of the pulse shape of two hydrogen lines in the spectrum radiated by a 4C35 hydrogen thyratron during a 1.0 microsecond pulse and analysed by the method described. The time scale runs from right to left in this figure. *a*) Line at 6563 Å emitted by atomic hydrogen. *b*) Line at 6032 Å belonging to the $3p^3\Pi - 3s^3\Sigma, 0-0$ band emitted by the hydrogen molecule.

The records are not typical of hydrogen thyratron pulses but were obtained in a case when there was a large current spike early in the 1 μsec pulse. The dynode voltage pulses used for scanning the 1 μsec light pulse duration had a measured length of 9×10^{-8} sec.

riate fixed setting of the delay circuit and scanning of the spectrum with the motor-driven monochromator ("time-resolved spectroscopy"). *Fig. 5* shows a portion of the hydrogen band spectrum, emitted 0.2 μsec after the initiation of the discharge, which was obtained in this fashion. *Fig. 6* is a series of spectra from a nitrogen discharge (duration of the electrical discharge 10^{-7} sec) demonstrating the

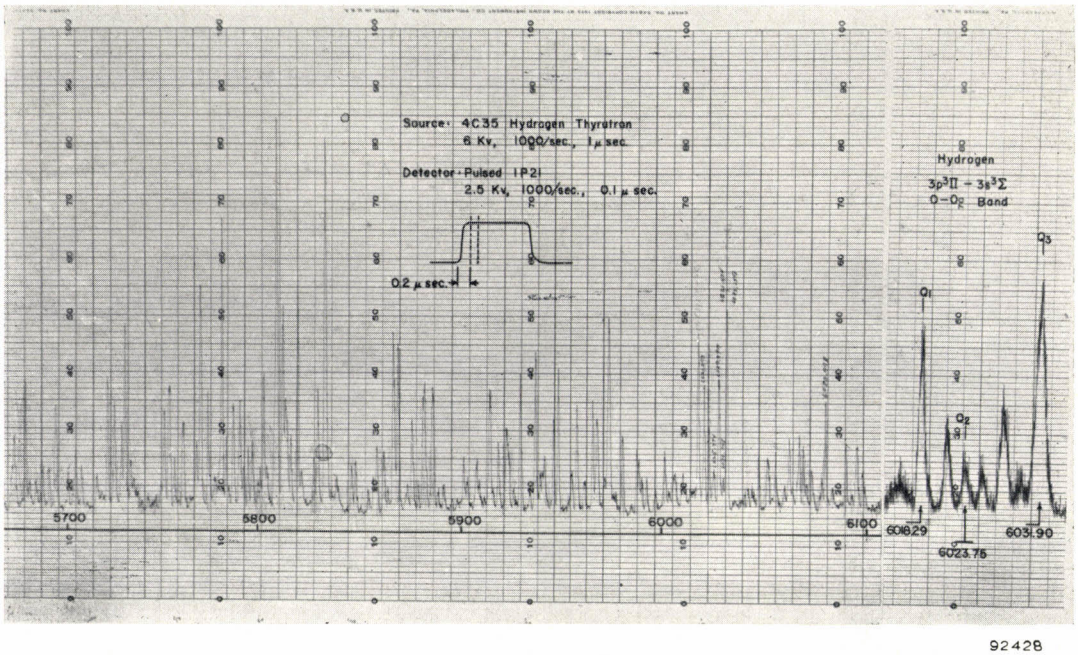


Fig. 5. Time-resolved spectroscopy: part of the hydrogen band spectrum emitted from a 4 C 35 hydrogen thyratron 0.2 μsec after initiation of a 1.0 μsec square current pulse. Thyratron peak voltage 6 kV, pulse repetition rate 1000 per sec. Right: narrow part of the spectrum recorded with greater dispersion.

rapid changes in the spectrum that may occur within a fraction of a microsecond. Such changes are due to differences in excitation and life of the molecules in the quantized energy levels participating in the emission of the various lines and bands.

Behavior of the photomultiplier tube under pulsed operation

An essential factor in these experiments is the internal *gain* of the photomultiplier tube. Much higher gains can be achieved under pulsed operation than with D.C. voltages on the tube, owing to the fact that higher voltages are permissible. This remarkable fact has been known and made use of for some time [7]). The oscillograms of single light flashes in fig. 1 in fact were obtained with a pulsed 1 P 21 photomultiplier tube at a voltage of 1500 V, since the gain of this tube under normal operation (maximum D.C. voltage 1230 V) would not have been sufficient to render visible each separate photo electron.

The gain was determined and its variation with voltage studied by means of a series of pictures similar to fig. 1, made under constant illumination. The average number $\bar{n}_p$ of photoelectrons per pulse was counted and the gain g calculated from the photomultiplier pulse repetition rate f and the measured average collector current $\bar{I}_p$, using the equation:

$$\bar{I}_p = \bar{n}_p f g e, \quad \ldots\ldots \quad (1)$$

e being the electronic charge. Within the admissible D.C. voltage range, the gains under D.C. and pulsed conditions were found to be equal. The gain under pulsed operation continued to grow as the voltage was increased, until a value of 0.8×10^9 was reached at 2500 volts (total voltage on the tube, i.e. between cathode and collector, the dynodes being supplied with equally spaced intermediate voltages taken from a potentiometer; cf. fig. 10). This is illustrated by *fig. 7*.

Incidentally, the oscillograms of fig. 1 neatly illustrate the effective shutter action of the pulsed photomultiplier tube: no signal whatever appears before or after the pulse. Practically all the pips occurring during the pulse are due to photoelectrons:

[7]) R. F. Post and N. S. Shiren, Performance of pulsed photomultiplier scintillation counters, Phys. Rev. **78**, 81, 1950. S. Singer, L. K. Neher and R. A. Ruehle, Pulsed photomultipliers for fast scintillation counting, Rev. sci. Instr. **27**, 40-43, 1956 (No. 1).

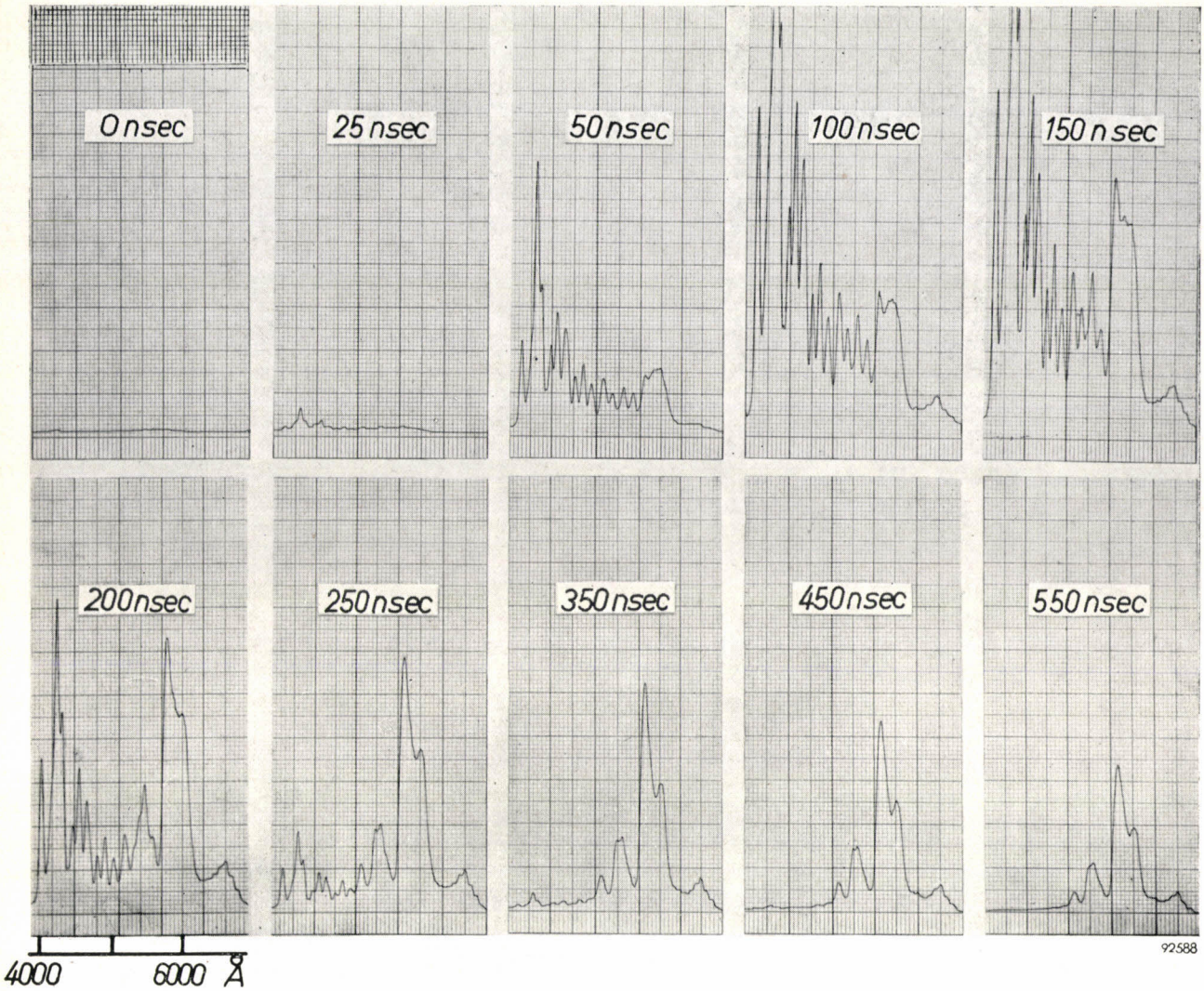


Fig. 6. Time-resolved spectroscopy: a series of spectra obtained from a pulsed nitrogen discharge at successive instants during the pulse, from 0 up to 550 nanoseconds (1 nsec = 10^{-9} sec) after the initiation of the discharge.

The narrow peaks on the left of each spectrum ($\lambda < 5000$ Å) are part of the "2nd Positive" bands of N_2. The broad peaks on the right ($\lambda > 5000$ Å) are groups of unresolved "1st Positive" bands. A sharp cut-off filter, Corning 3389, was used in these runs to eliminate the strong 2nd Positive bands at $\lambda < 4000$ Å, lest their second order spectrum should obscure the much weaker 1st Positive bands. Separate investigation of these bands showed that all 2nd Positive bands had about the same time variation.

Note that the 2nd Pos. bands attain their peak intensity in 100-150 nsec and have practically disappeared at 250 nsec. The 1st Pos. bands appear later (peak in 150-200 nsec) and decay much more slowly. This is explained by the fact that the lower quantum level of the 2nd Pos. bands transition is the upper quantum level for the 1st Pos. bands transition, and this level apparently has a much longer half-life ($\sim 10^{-6}$ sec) than the upper level of the 2nd Pos. bands ($\sim 10^{-7}$ sec) (although other factors may contribute to the long decay time of the 1st Pos. bands, such as absorption from the lower (metastable) level of this transition with subsequent re-emission, or a populating of the (triplet) upper level of this transition from some metastable singlet level as proposed by R. W. Nicholls, J. chem. Phys. **20**, 1040, 1952). Direct experimental data on the half-lives of these levels have not been previously reported.

In similar runs, but under greater spectral dispersion, the individual 1st Pos. bands were resolved. The relative intensities of these bands did not vary greatly during the decay and were approximately equal to those obtained in a normal D.C. glow discharge in nitrogen. This point distinguishes the long decay which we observed ($\sim 10^{-6}$ sec) from the well-known nitrogen afterglow (decay of the order of 1 second), since in the latter case the relative intensities of the bands are entirely different.

when the light is blocked, they do not appear, except for an occasional isolated pip, about one every ten phototube pulses, which permits the conclusion that such "dark pips" are due to thermal electrons emitted at the cathode.

The oscillograms of *fig. 8* will provide an explanation for the higher voltages admissibles under pulsed operation and give an indication of the limits imposed. These pictures were obtained with 3 μsec pulse voltages higher than that used for fig. 1 and with *no light* on the tube. At a voltage little above 2500 V, pips of low amplitude began to appear, at first very erratically and only near the end of the 3 μsec pulse. As the voltage was raised these "noise" pips became more regular, greater in amplitude and number, and appeared earlier in the pulse. The interpretation is that these pips are due to positive ions from residual gas in the tube striking the dynode

surfaces and initiating electron avalanches. This gas-generated noise apparently requires several microseconds or even longer to develop at voltages below 2 kV, but less time as the voltage is increased. Even at voltages of 4 kV, however, it is seen that the tube is relatively free of the noise during the first few tenths of a microsecond, as would be expected from calculated transit times of positive ions. It is this delay in the development of gas noise that allows these tubes to be used at very high gains under 10^{-7} sec pulse conditions.

The noise pattern as shown in fig. 8 (obtained for 3 μsec pulses) is unaffected by repetition of the pulses up to rates of at least 1000 per second. Evidently, the tube will recover from the severe gas ionization in less than one millisecond.

Fig. 8 shows the behaviour of the pulsed tube with no light. In actual use, the light falling on the tube will enhance the gas noise breakdown and will evoke the appearance of noise pips at voltages much lower than those used in fig. 8. This may affect the linearity of the relative intensity measurements to some extent. It was found, however, that the magnitude of the noise can be greatly reduced by painting the outside of the photomultiplier tube with finely dispersed graphite and applying to this conducting layer either the cathode pulse voltage or a D.C. voltage equal to its peak. This procedure probably reduces the electric field in the tube and thereby the gas ionization.

It is important to note that a flash of light falling on the pulsed tube in the *beginning* of its "on" time will evoke noise pips continuing during the whole

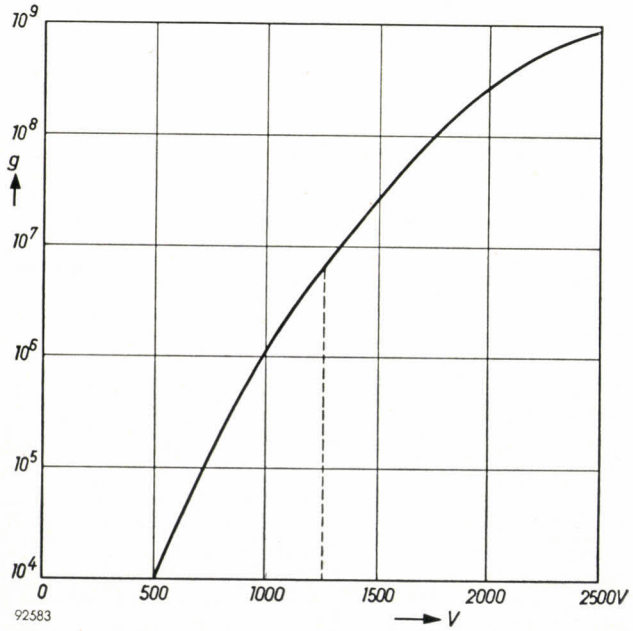


Fig. 7. Gain g obtained when actuating a 1 P 21 photomultiplier tube with a 10^{-7} sec pulse, plotted against the peak voltage applied between cathode and collector.

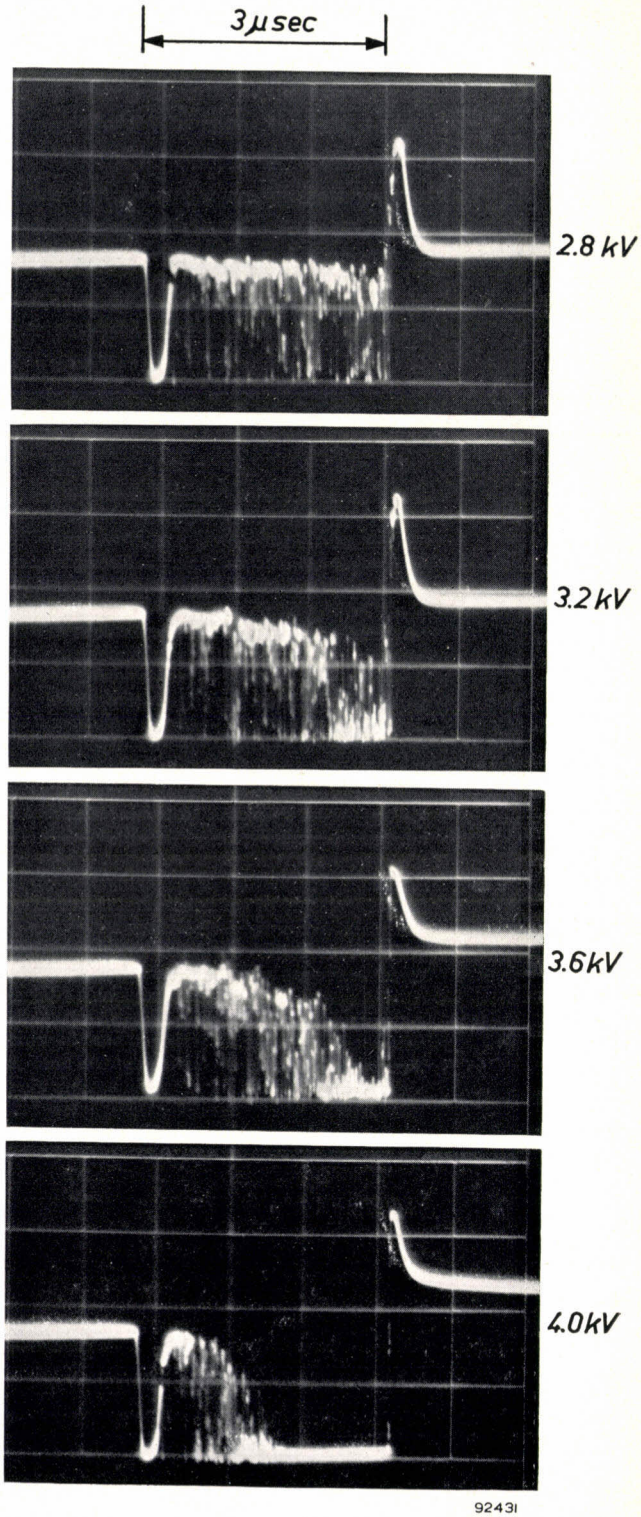


Fig. 8. Oscilloscope traces of the output of a photomultiplier tube during a single 3 μsec pulse, with *no illumination*, at high pulse voltages (peak value indicated at the right). The numerous pips probably are due to positive ions from residual gas in the tube (gas noise). The density of pips at the highest voltages becomes so large towards the end of the pulse as to cause a continuous trace.

"on" time (after-pulsing). No noise pips appear, however, if the light flash is extinguished *before* the start of the photomultiplier pulse, even if it precedes the latter by as little as 10^{-7} sec. This is

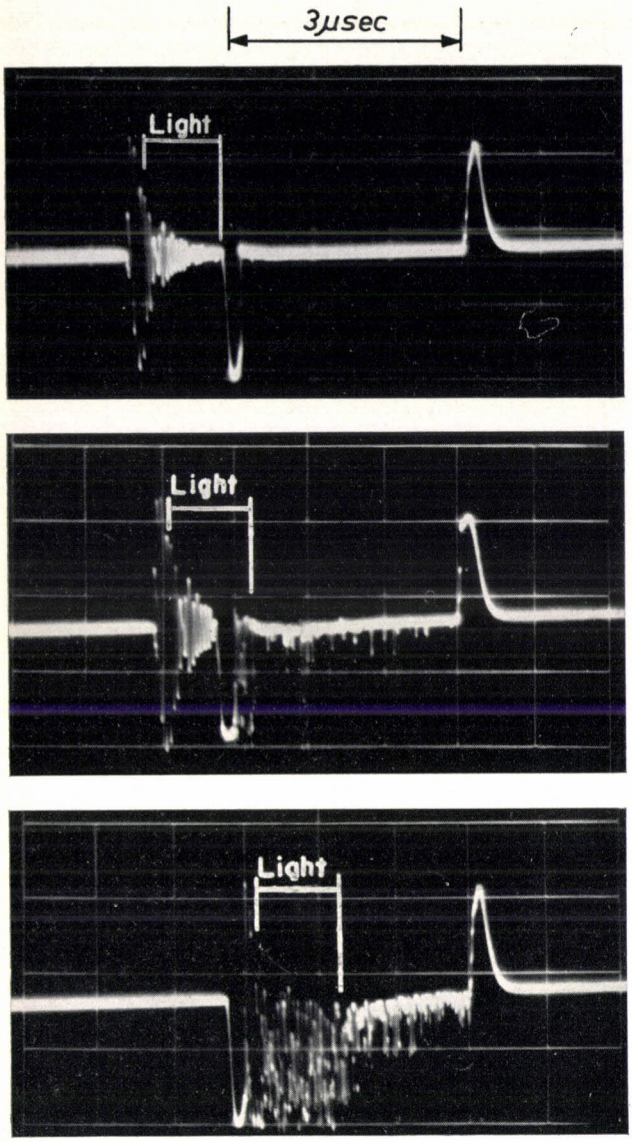


Fig. 9. Same as in fig. 8, at a pulse voltage of 2.0 kV, with a 1 μsec light flash falling on the phototube just prior to or in the beginning of the pulse. No afterpulses whatever are recorded in the first case.

The initiation of the light flash becomes visible in the oscilloscope trace as a damped oscillation picked-up electrically by the oscilloscope from the 20 kV pulse feeding the light source. It is a remarkable fact that the response of the pulsed photomultiplier itself is quite unaffected by such pick-up phenomena. This enables the leading edge of light flashes (build-up phenomena in electrical discharges) to be effectively studied by the stroboscopic method decribed.

demonstrated by the series of oscillograms reproduced in *fig. 9*, showing the signal of a phototube during a 3 μsec pulse, on exposure to a 1 μsec light flash just preceding or overlapping the phototube pulse. Because of this remarkable behaviour of the phototube, the pulsing technique was used in this laboratory in several instances to detect very weak light following an intense light flash, such as in the study of weak afterglows and excitations caused by absorbed light and also in an optical ranging system. In the latter, a bright flash of light is directed at a distant object and the reflected pulse is detected by a pulsed photomultiplier. The weak reflected pulse, received only 0.5 μsec after the initial flash, was detected against the unavoidable local back-scatter which, at the phototube, may be many times greater in intensity. Experiments showed that an intensity ratio of 10^5 between back-scatter and reflected flash was permissible.

A final, not fundamentally but practically important remark concerning the pulsed operation of photomultiplier tubes should be made: the need for light-tight shielding and low level room lighting is much relaxed because of the low duty factor. Further the danger of damage of the tube due to exposure to room light is less because the current is not allowed to assume excessive mean values even with the extremely high gain obtained at high voltages.

Circuits for pulsing the photomultiplier tube

In order to accomplish the required pulsing and phasing, several forms of circuitry have been tried. *Fig. 10* shows one form, using only standard tubes and components. The upper portion of the circuit is for triggering the light source while the lower portion generates the phototube voltage pulse, usually of 10^{-7} sec duration, with peak adjustable from 0 to 5000 V. The common trigger generator T of both portions was designed to supply 1-10000 pulses per second. The phase relation between the two output pulses of the circuits is controlled by the potentiometer *Pot* and the step delay line D. The duration of the phototube voltage pulse is determined by the length of the coaxial cable C, used as a pulse forming network.

For the stroboscopic operation with an "on" time of 10^{-7} sec, it was necessary to make the relative time jitter between the light pulse and the phototube pulse less than 10^{-8} sec. In addition this accuracy had to be maintained when the phase between the two pulses was adjusted over a 5 μsec range. This was the most difficult experimental problem encountered and was accomplished only after careful decoupling, shielding and power supply regulation. Later modification of the circuit incorporated continuously variable time delay networks for the phase control and this greatly reduced the jitter.

The phototube pulse was capacity-coupled to the dynode voltage divider B, consisting of ten 100 ohm carbon resistors. This low impedance is required to allow the dynode voltages to rise and fall in times short compared to the 10^{-7} sec pulse duration. It was found that the interelectrode capacitance of the

normal 11-pin tube socket of the photomultiplier tube caused too large a pulse round-off. Therefore the voltage divider was connected directly to the base pins of the phototube, as shown in fig. 3. Removing the base of the tube did not give any further improvement.

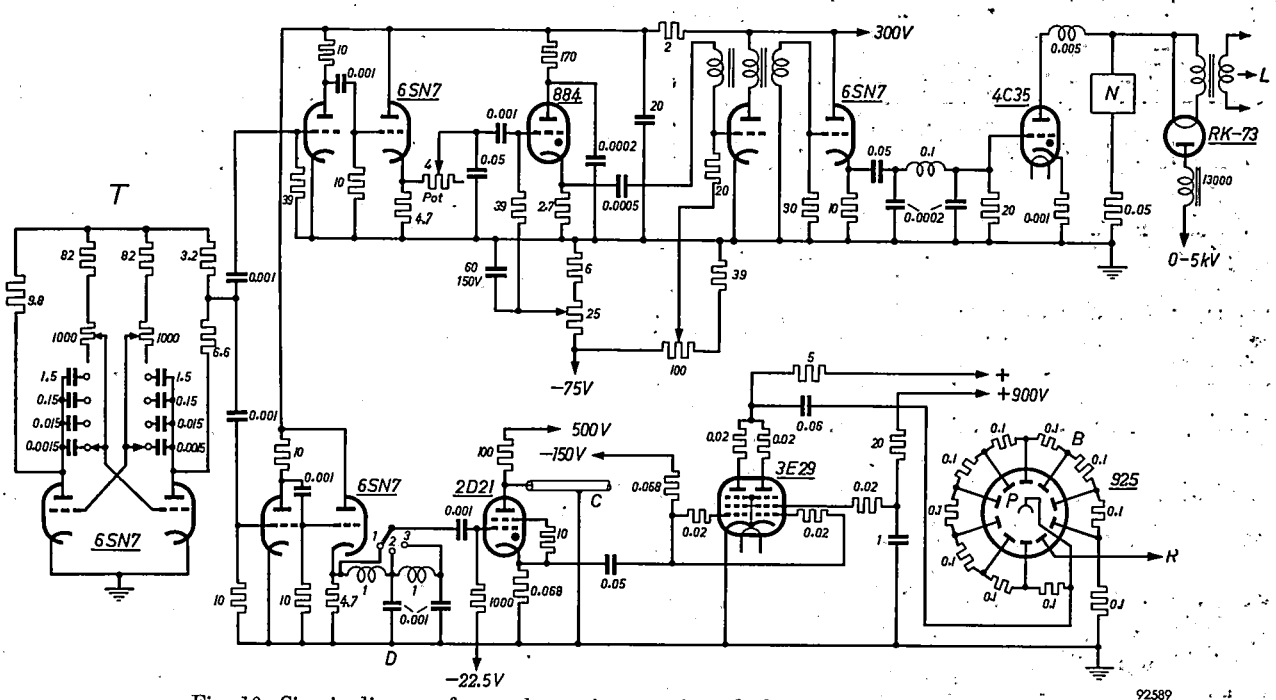


Fig. 10. Circuit diagram for stroboscopic operation of photomultiplier tube. The upper portion produces the pulse fed to the light source (L), the lower portion supplies the required voltage pulse (peak 0-5000 V) to the photomultiplier tube P. T common trigger generator; the potentiometer *Pot* and step delay line D control the phasing of light flash and phototube pulse, the coaxial cable C determines the duration of the phototube pulse; the pulse-forming network N controls the form and duration of the light flash. B is the voltage divider for the dynodes. The recording microammeter is connected at R.

Several experiments were made in which the voltage applied to only one or two dynodes of the photomultiplier tube was pulsed while the remaining dynodes were maintained at their normal D.C. operating potential. In such an operation one of the dynodes would act as a control grid; the complete cut-off required for the shutter action, however, was not obtained. Some signal was recorded even though the control dynode voltage was allowed to swing several hundred volts in either direction.

The current meter used was a commercial electronic microammeter suitably modified to give full-scale readings at currents down to 10^{-9} amperes, and an output for connection to a 0-10 mV strip chart recorder. Across the input of the meter was placed a suitable capacitor whose capacitance, together with the input impedance of the meter, produced the desired RC time constant, for the integration of the signal in successive pulses.

Sensitivity and accuracy of the measurements

It should be realized that according to eq. (1) an average phototube current of 10^{-9} A with a gain $g = 10^9$ and a pulse repetition rate f of say 1000 per sec corresponds to $\bar{n}_p \approx .0.006$, i.e. less than 1 photoelectron per 100 phototube "on" times! Such extremely low intensity signals can be detected and reasonably accurately measured owing to two features of the stroboscopic operation, qualitatively already implied in the description given above: 1) Parasitic current contributions, mainly from thermal electrons, occurring at a rate of about 10^5 per second at 300 °K, are effectively eliminated by the gating action of the phototube: the average number $\bar{n}_t$ of thermal electrons per pulse is about 0.01 (again 1 electron per 100 phototube "on" times). 2) The fluctuations of the measured current are greatly reduced since the rate of the contributing photoelectrons is averaged over a very long RC time. The "signal-to-noise ratio". S/N of the phototube collector current is given by the expression [8]):

$$\frac{S}{N} \approx \sqrt{\frac{\overline{g^2}}{\bar{g}^2} \frac{\bar{n}_p}{\bar{n}_p + \bar{n}_t} \cdot 2\bar{n}_p f RC}, \quad . \quad . \quad . \quad (2)$$

taking into account the fluctuations of the gain due to random fluctuations in the secondary emission. Assuming a value of 0.75 for $\overline{g^2}/\bar{g}^2$ (determined for a

[8]) From an unpublished analysis by F. K. du Pré of this laboratory.

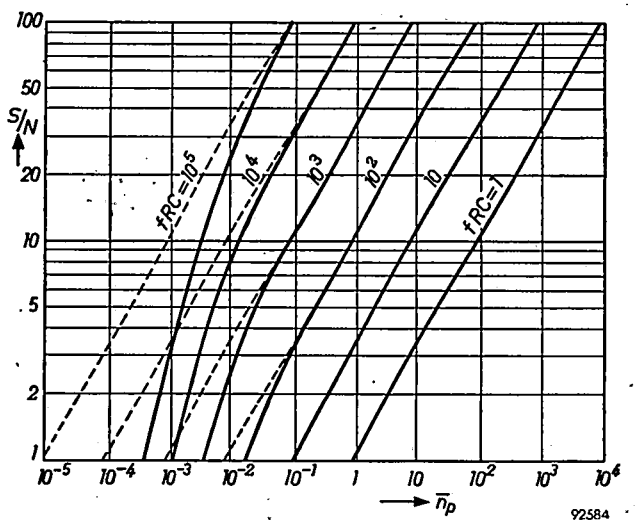


Fig. 11. Signal-to-noise ratio S/N of the photomultiplier-tube collector current, plotted against the average number $\bar{n}_p$ of primary photoelectrons per photomultiplier pulse, for different pulse repetition rates f and integration time constants RC (see eq. 2). The dashed curves would be obtained if no thermal electrons were emitted by the photocathode.

particular RCA 1 P 21 photomultiplier tube by measuring the heights of a large number of pips due to individual electrons in photographs of the type of fig. 1), S/N is plotted against $\bar{n}_p$ in *fig. 11*. It can be seen that a signal-to-noise ratio $S/N = 4$ can be obtained with light flashes producing as little as 1 photoelectron in 1000 flashes, provided the maximum repetition rate $f = 10^4$ per sec and an RC time of 10 seconds are used for the stroboscopic operation.

Since the actual "on" time of the photomultiplier tube is equal to the duration of the dynode voltage pulse minus the total transit time of an electron avalanche through the tube (a few 10^{-8} sec), the *spread* in the transit time, which may amount to a few 10^{-9} sec, will be fully felt in the "on" time and will therefore ultimately limit the time resolution of which the stroboscopic scanning method is capable. A practical limit on the resolution is imposed by the jitter caused by the circuitry as mentioned above. In fig. 4 the resolution obtained in scanning a periodic 1 μsec light flash by means of 10^{-7} sec phototube pulses is seen to be better than 3×10^{-8} sec (half-value width of the first Hα-peak in fig. 4*a*).

Summary. The time variation of the light intensity in weak light flashes can be analyzed by using a photomultiplier tube supplied with pulsed dynode voltages. The pulsed tube acts as an effective light shutter, providing an "on" time as short as 10^{-7} sec if desired. By periodic repetition of the light flash and synchronization with the photomultiplier tube pulses in a constant phase relation, separate instants of the light flash can be isolated and the instantaneous intensity accurately measured by integrating the phototube collector current over a long period. Circuits are described which permit repetition rates up to 10 000 per sec and integration periods of 10 sec or even longer. The 10^{-7} sec phototube pulse can be phased through the complete light flash (of say 10^{-6} sec duration), providing a stroboscopic analysis of the flash. The dynode voltages in the pulsed operation of the phototube can be raised considerably above those admissible in normal D.C. operation, resulting in a much higher gain (up to 10^9) and permitting the measurement of extremely low light intensities, e.g. in narrow spectral regions of the total light emitted. Examples of "time-resolved spectograms" obtained by this method are presented in this article and some details concerning the pulsed operation of the photo-multiplier tube and the accuracy of the measurements are discussed. The signal-to-noise ratio can be high even with signals producing a fraction of one photoelectron per pulse, since the gating action of the pulsed phototube effectively suppresses thermionic and other noise current contributions. The time resolution can be made better than 3×10^{-8} sec by effectively reducing the time jitter between light pulses and phototube pulses.

APPLICATIONS OF THE INTERFERENCE OF LIGHT IN THIN FILMS

by P. M. van ALPHEN.

535.417:621.397.2

The phenomenon of the interference of light in thin films has been put to use for some time in optical instruments, e.g. prismatic binoculars, and in recent years also in colour television. In view of some articles on the latter subject, which are to appear shortly in this Review, the article below gives a short discussion of the theory and of the various applications of interference at thin films.

Probably everyone, as a child, has been fascinated by the phenomenon of optical interference: the multi-coloured patch of oil on wet asphalt and the beautiful colours of a soap bubble in the sun will be familiar to everybody. Centuries ago these colours provoked curiosity; Newton and many others investigated these phenomena.

For many years now, interest in this interference phenomenon has outgrown its purely scientific aspects: it is also of practical value. We shall discuss here some aspects of this technical development, thereby, it may be, illuminating the fascination of a soap bubble.

Optical interference in a thin transparent film

When light is reflected from a thin transparent film, interference occurs and, as a result, certain colour effects can be observed. To make this clear, let us consider *fig. 1*, in which *1* represents a very thin film of water spread over a glass plate *2*. If a beam of white light falls almost perpendicularly

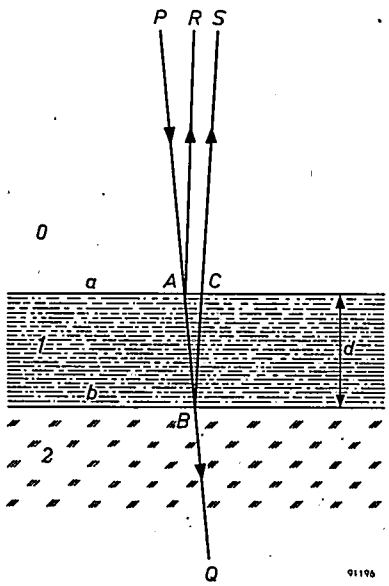


Fig. 1. Reflection, for almost normal incidence, of ray PA (in air, medium *0*) at a film of water (medium *1*) on glass (medium *2*). $PABQ$ and BCS are assumed to be straight lines.

on it, some portion of the light will be reflected from each of the two boundary surfaces, a and b. The reflected rays are parallel ($AR//CS$). Since the light reflected at B has travelled a longer path than that reflected at A, and since both rays originate from the same source, a phase difference occurs between the two waves and interference can take place. If the phase difference corresponds to a whole number of wavelengths, the waves reinforce each other and the intensity of the reflected light is increased; if the phase difference represents an odd number of half wavelengths, then the waves oppose one another and the intensity of the reflected light is decreased, perhaps to zero. In the first case we have a *high-reflection* film, in the second, a *low-reflection* film. Since the phenomena are dependent on the wavelength, increased reflection for one colour can coincide with decreased reflection for another.

The magnitude of the phase difference is easily found. For normal incidence the path difference is twice the thickness of the film d. The wavelength of light in water is λ/n, where λ is the wavelength of light in air and n the refractive index of water. The phase difference in this case is thus $(2nd/\lambda)\times 2\pi$ radians. If the rays do not strike at right angles, but at an angle α to the normal to the film (*fig. 2*), the phase difference becomes:

$$\frac{2nd\cos\vartheta}{\lambda}\times 2\pi \text{ radians}, \quad \ldots\ldots \quad (1)$$

where ϑ is the angle which the refracted ray AB makes with the normal ($\sin\alpha/\sin\vartheta = n$).

Since the path in water ABC is equal to $2d/\cos\vartheta$, one might suppose that in equation (1) the factor $\cos\vartheta$ should be in the denominator. Therefore we give here in brief the derivation of (1).

The "optical path" of the ray ABC in water of refractive index n is $2nd/\cos\vartheta$. The light reflected at A must travel a distance AC' before it reaches the wave front of C. Now $AC' = AC\sin\alpha$, $AC = 2d\tan\vartheta$ and $\sin\alpha/\sin\vartheta = n$, thus

$$AC' = 2d\tan\vartheta\sin\alpha = 2nd\frac{\sin^2\vartheta}{\cos\vartheta}.$$

The optical path difference is therefore:

$$2nd\left(\frac{1}{\cos\vartheta}-\frac{\sin^2\vartheta}{\cos\vartheta}\right)=2nd\cos\vartheta,$$

as stated in (1).

The phase difference is thus less for obliquely incident rays than for those with normal incidence.

Besides this phase difference, there is sometimes a phase jump as well, in cases where the reflection takes place at an optically denser medium, analagous to the phase jump occurring at the end of vibrating strings, organ pipes, Lecher lines, waveguides, etc.

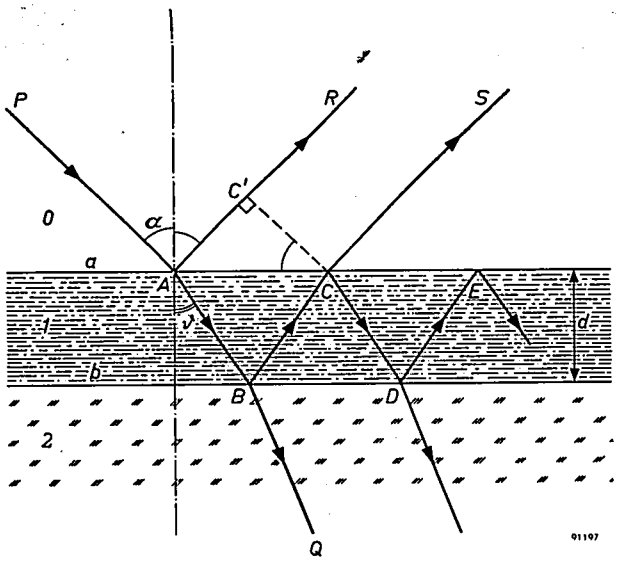


Fig. 2. As in fig. 1 but now for oblique incidence. Some of the multiple reflections are shown.

For normal incidence, the phase jump is π radians. In the example shown in fig. 1, both reflections take place at optically denser media (water is denser than air, glass denser than water). Both at A and at B, for normal incidence, a phase jump of π radians occurs, so that the phase difference is not affected. However, if one is dealing with a free film, such as a soap bubble, or if layer *1* in fig. 1 is of a material which is optically denser than glass, then only the first reflection takes place at an optically denser medium, while the second occurs at an optically less dense medium and no phase jump occurs for the latter. For normal incidence the condition for maximum reflection is then:

$$\frac{2nd}{\lambda}+\tfrac{1}{2}=k, \quad \ldots\ldots \quad (2a)$$

and that for minimum reflected light:

$$\frac{2nd}{\lambda}+\tfrac{1}{2}=k+\tfrac{1}{2}, \quad \ldots\ldots \quad (2b)$$

where k is an integer.

Reflection coefficient of a boundary surface and of a layer.

Let us now consider the reflection coefficient r, for the case of a ray travelling in one medium (e.g. air) and being reflected at the boundary surface of a denser medium (e.g. glass). r is dependent on the refractive index n from the one medium to the other, the angle of incidence and the direction of polarisation, and is given by the formulae of Fresnel. At normal incidence these formulae reduce to

$$r=\left(\frac{n-1}{n+1}\right)^2. \quad \ldots\ldots \quad (3)$$

For air-glass ($n = 1.5$) we find: $r = 0.04$, and for air-zinc sulphide ($n = 2.4$): $r = 0.17$.

Equation (3) can easily be deduced from the fact that according to Maxwell's theory the tangential components of both the electrical field strength $\boldsymbol{E}$ and the magnetic field strength $\boldsymbol{H}$ at the boundary between two media are equal (*fig. 3*):

$$\left.\begin{aligned}E_1-E_1'&=E_2,\\ H_1+H_1'&=H_2,\end{aligned}\right\}, \quad \ldots\ldots \quad (4)$$

there being no other components for normal incidence. (Suffix 1 refers to one medium, suffix 2 to the other and the accent to the reflected ray. There is a negative sign in one equation and a positive sign in the other because the incident and reflected rays move in opposite directions: thus one of the vectors, e.g. the $\boldsymbol{E}$-vector, is reversed in direction.)

Now $H_1=\sqrt{\varepsilon_1}E_1$, $H_1'=\sqrt{\varepsilon_1}E_1'$, $H_2=\sqrt{\varepsilon_2}E_2$ and $n=\sqrt{\varepsilon_2/\varepsilon_1}$ (ε_1 and ε_2 are the dielectric constants), so that instead of (4) we can write:

$$E_1-E_1'=E_2,$$
$$E_1+E_1'=nE_2.$$

By subtraction and addition we obtain:

$$E_1'=\tfrac{1}{2}(n-1)E_2,$$
$$E_1=\tfrac{1}{2}(n+1)E_2,$$

thus

$$\frac{E_1'}{E_1}=\frac{n-1}{n+1}.$$

The reflection coefficient r is the square of this amplitude ratio, thus

$$r=\left(\frac{n-1}{n+1}\right)^2. \quad \ldots\ldots \quad (3)$$

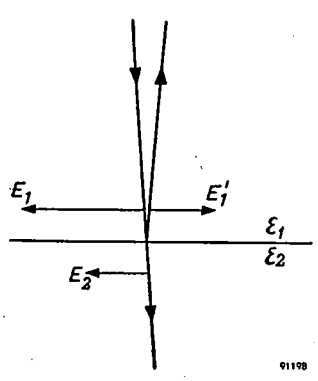


Fig. 3. The almost normally incident ray and the reflected and the refracted rays at the boundary of two media of dielectric constants ε_1 and ε_2. The tangential components of the electric field are E_1, E_1' and E_2 respectively. The tangential components of the magnetic field are perpendicular to the plane of the drawing and all point in the same direction.

If we are dealing not with one boundary surface, but with two (or more), as in the case of the water layer (fig. 2) or a soap film, then for accurate results, account must be taken of repeated reflections: only a portion (CS) of the ray BC in fig. 2 leaves the film, while the remainder is reflected as CD; of this a further portion is reflected as DE, and so on. At all these reflections a phase jump may or may not take place, according as the refractive index of the film is smaller or larger than that of the adjoining medium. Taking all this into account, we find for the reflection coefficient R of the layer:

$$R = \frac{\varrho_a^2 + 2\varrho_a\varrho_b \cos 2\delta + \varrho_b^2}{1 + 2\varrho_a\varrho_b \cos 2\delta + \varrho_a^2\varrho_b^2}. \quad . \quad . \quad (5)$$

In this expression

$$\delta = \frac{2\pi}{\lambda} n_1 d \cos\vartheta, \quad . \quad . \quad . \quad . \quad (6)$$

and ϱ_a^2 and ϱ_b^2 are the Fresnel reflection coefficients at the boundary surfaces a and b respectively. For normal incidence ϱ_a and ϱ_b take the simple form:

$$\varrho_a = \frac{n_0 - n_1}{n_0 + n_1}, \qquad \varrho_b = \frac{n_1 - n_2}{n_1 + n_2}.$$

n_0, n_1 and n_2 are the refractive indices (with respect to vacuum, or air) of the media 0 [1]), 1 and 2. Whether or not a phase shift occurs depends on the sign of ϱ_a and ϱ_b.

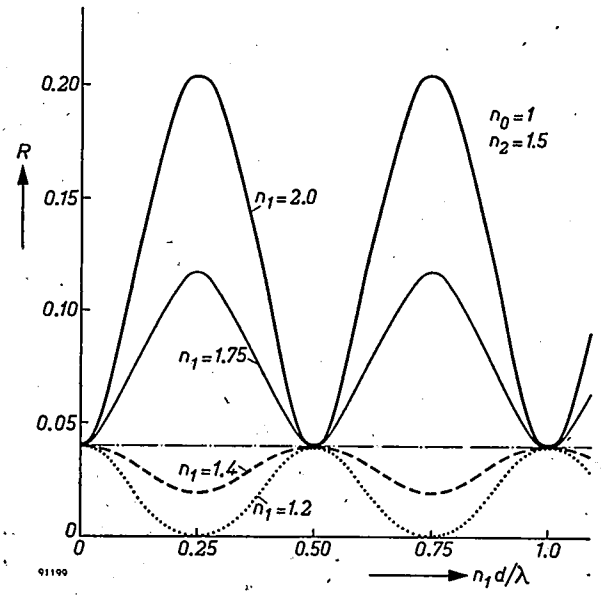


Fig. 4. Reflection coefficient R of a thin film (thickness d, refractive index n_1) as function of $n_1 d/\lambda$, for various values of n_1. The refractive indices n_0 and n_2 of the adjacent media are 1 and 1.5 respectively.

[1]) By putting $n_0 = 1$ in the expression for ϱ_a, we find for ϱ_a^2 the same as equation (3), where the medium 0 was assumed to be air.

Substituting the expressions for ϱ_a and ϱ_b in (5), we find for R at normal incidence:

$$R = \frac{(n_0^2+n_1^2)(n_1^2+n_2^2) - 4n_0n_1^2n_2 + (n_0^2-n_1^2)(n_1^2-n_2^2)\cos 2\delta}{(n_0^2+n_1^2)(n_1^2+n_2^2) + 4n_0n_1^2n_2 + (n_0^2-n_1^2)(n_1^2-n_2^2)\cos 2\delta}. \quad (7)$$

The thickness of the film d and the wavelength λ occur as d/λ in the angle δ, see (6); the fact that the refractive indices depend to some extent on λ (dispersion) has been neglected.

Fig. 4 shows R as a function of $n_1 d/\lambda$ drawn for several values of n_1, with $n_0 = 1$ (air) and $n_2 = 1.5$ (glass). The full lines ($n_1 > n_2$, no phase jump at boundary b) show high reflection, the broken and dotted lines ($n_1 < n_2$, phase shift at the boundary) show low reflection. Extreme values of R occur when $n_1 d = (2k+1)\lambda/4$ and when $n_1 d = 2k\lambda/4$. The magnitude of the extreme value and its nature (maximum or minimum) are given below:

Extreme value of R	$n_2 > n_1 > n_0$ (broken lines fig. 4)	$n_2 < n_1 > n_0$ (full lines fig. 4)
$\left(\frac{n_1^2 - n_0n_2}{n_1^2 + n_0n_2}\right)^2$	minimum at $n_1 d = (2k+1)\lambda/4$	maximum at $n_1 d = (2k+1)\lambda/4$
$\left(\frac{n_2 - n_0}{n_2 + n_0}\right)^2$	maximum at $n_1 d = 2k\lambda/4$	minimum at $n_1 d = 2k\lambda/4$

Low-reflection films

From the table above can be seen that the minima in the broken curves become zero if $n_1 = \sqrt{n_0 n_2}$ (i.e. if $n_1 = \sqrt{n_2}$ when medium 0 is air), and the minima occur at a film thickness $d = \frac{1}{4}\lambda/n_1$, $\frac{3}{4}\lambda/n_1$, and so on. The reflection from a glass surface could thus be reduced by the application of a film of a material for which $n_1 = \sqrt{1.5} = 1.23$, so reducing the loss of light by reflection in optical instruments. Materials with such a small refractive index, however, are unknown. In practice, one uses materials with the smallest possible refractive index, such as $NaAlF_4$ (cryolite) for which $n_1 = 1.35$, or MgF_2 for which $n_1 = 1.38$, applying them in such a porous form that the effective refractive index is about 1.30. The reflection which, for untreated glass, is 4.2%, is then reduced to 0.25% for the wavelength λ for which $n_1 d = \frac{1}{4}\lambda$. For light of wavelength $\lambda' = \frac{1}{2}\lambda$ the reflection is of course a maximum, but its magnitude is $(n_2-1)^2/(n_2+1)^2$, which is the value for reflection from the untreated glass surface; in other words, the layer does not increase the reflection for any wavelength (see

fig. 4). The thickness of the film is usually chosen so that the reflection reduction occurs in the region of the green; the glass then still reflects some red and blue, which gives it a violet appearance. An important point is that the layer absorbs very little. In instruments with many glass surfaces, e.g. prismatic binoculars and multi-component lenses, this reduction of the reflection can considerably improve the contrast.

By applying two or three films of different materials to the glass, the reflection can be brought to a minimum for two or three different wavelengths and so greatly reduced over the whole visible range. The effect is then rather strongly dependent on the angle of incidence. If the thicknesses of the films are chosen for normal incidence, then the path differences will be smaller for other angles of incidence (factor $\cos \vartheta$ in (1)) — in varying degree in the different layers — which will reduce the effect.

High-reflection films

A high-reflection film can be of use, for example, to improve the reflection from a metal mirror. By coating aluminium (reflection coefficient 89%) with a $\frac{1}{4}\lambda$-film of SiO and a $\frac{1}{4}\lambda$-film of TiO_2, it is possible to give it a reflection coefficient of 95%.

Even more interesting is its use for semi-reflecting mirrors, which show negligibly small absorption. A semi-reflecting mirror in the form of a thin transparent film of metal always produces considerable absorption; silver, for instance, reflects 30%, transmits 30% and absorbs 40%. If we apply to glass a material for which $n_1 = 2.7$, then according to the first formula in the table above, the maximum of R is equal to 43%; owing to the negligible absorption, the transmission is thus 57%. In this way one obtains a semi-reflecting mirror which is practically free from losses. TiO_2, Sb_2S_3 and ZnS are suitable for use as the material for the film. If, instead of applying the film directly to the glass, an intermediate layer of thickness $\frac{1}{4}\lambda/n_2$ and smaller refractive index are used, the reflection can be increased to 52%.

Colour-separating (dichroic) mirrors

Up to now, we have mainly been concerned with light of a constant wavelength λ. We shall now consider what happens when λ is variable and the film thickness d constant. If we once more neglect dispersion, we can read off the reflection coefficient R of a film as a function of λ^{-1} from fig. 4. We see that in the case where $n_2 < n_1 > n_0$, light of wavelengths $\lambda_{max} = 4n_1d/(2k+1)$ suffers maximum reflection and light of wavelengths $\lambda_{min} = 4n_1d/2k$ minimum reflection, i.e. maximum transmission. Partial separation of the colours thus occurs.

This effect can be enhanced in the same way that the purification of materials by distillation or recrystallisation can be improved, namely by repeating the process, in this case by allowing the light to pass through another thin layer. For this purpose a number of thin layers of high refractive index must be applied one after the other. It is clear that these layers must remain separate from each other, or the necessary boundary surfaces would disappear and one would obtain a single thick layer. The separation is achieved by means of a material of low refractive index. The technique of applying such films will be dealt with presently.

There is, however, one more condition which must be fulfilled. The rays transmitted and reflected by the various layers will exhibit relative phase differences; in combining a number of layers, therefore, one must ensure that the maximum from one layer reinforces that from the following layer and does not cancel it. The layers must thus lie at a very definite distance from one another, that is to say, the separating layers of low refractive index must be of sufficient thickness. The separating layers are usually composed of alkali fluorides which, like the other layers, are applied by evaporation. These are essentially the same materials as are suitable for low-reflection films.

Since corresponding points on the reflecting layers must lie at an optical distance of $\frac{1}{2}\lambda$ (i.e. an actual distance of $\frac{1}{2}\lambda/n_1$) apart — the phase difference must be 2π, but the distance is travelled twice — and the layer itself has an optical thickness of $\frac{1}{4}\lambda$, the intermediate layers should also produce an optical separation of $\frac{1}{4}\lambda$. One arrives thus at a pile of layers of alternating high and low refractive index, all with an optical thickness of $\frac{1}{4}\lambda$.

By this combination of a number of layers the reflection as function of λ^{-1} takes on another form: the curve becomes steeper and secondary maxima appear in increasing numbers as the number of layers is increased. The steepness of the curve is easily understood when one remembers that with, say, seven layers even the first and last surfaces must still work together; small deviations in d or λ are additive and are more likely to produce appreciable phase differences when more layers are involved. By not making the layers all of exactly the same thickness, the secondary maxima can be weakened without much effect on the principal maximum.

The problem of calculating the interference in a multiple layer has much in common with that in

the optical grating where the light passes through a number of narrow slits placed close to each other. The multiple layer is a sort of "plane grating". The mathematical treatment is more difficult than that of the slit grating, in the first place because account must be taken of repeated reflections between the layers, and further because the intensity of the light falling on a given layer is different from that on the preceding layer, since a portion is reflected at each layer. Various methods, partially graphic, are known by which the theoretical curve can be calculated. *Fig. 5* shows the results of a calculation for a single layer of zinc sulphide and for three, five and seven layers, alternated with cryolite, with glass as the carrier. Experimentally determined curves generally differ from those calculated because the layers are not sufficiently homogeneous and not everywhere of exactly the same thickness, and because dispersion has been neglected. *Fig. 6* shows two measured curves relating to dichroic mirrors made in this laboratory.

Preparation

Before discussing the applications of dichroic mirrors, we shall say something about their preparation.

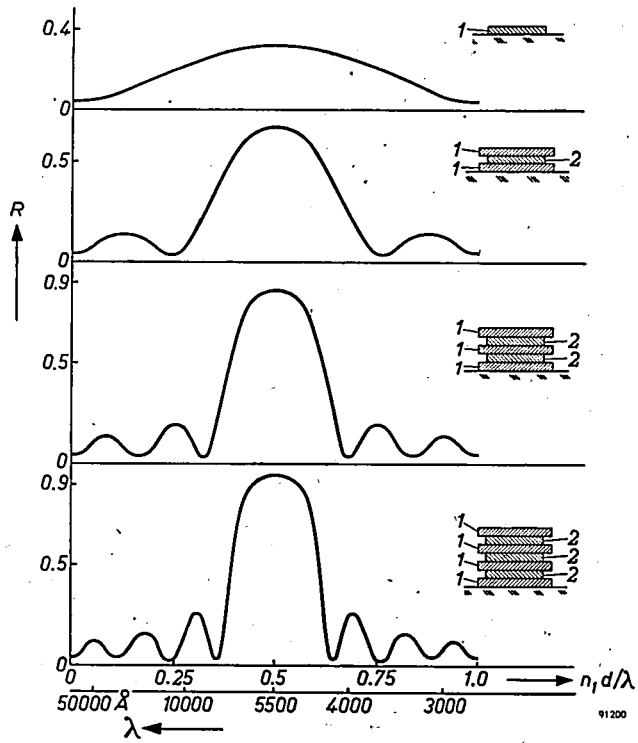


Fig. 5. Reflection coefficient R (calculated) for a single layer of zinc sulphide and for three-, five- and seven-fold layers on glass. The multiple layers consist of alternate layers of zinc sulphide *1* (thickness d, refractive index 2.4) and cryolite *2* (refractive index 1.3); the carrier substance is glass (refractive index 1.5).

As was shown above, one requires layers of optical thickness $\frac{1}{4}\lambda$ or an odd multiple of this quantity, thus of the order of 10^{-4} mm. Apart from mechanical methods such as rolling (various metals), beating (gold) and blowing (glass, soap bubbles), other methods were known of old by which thin films could be produced, such as etching, oxidizing and "burning" (platinum). In the last decades the

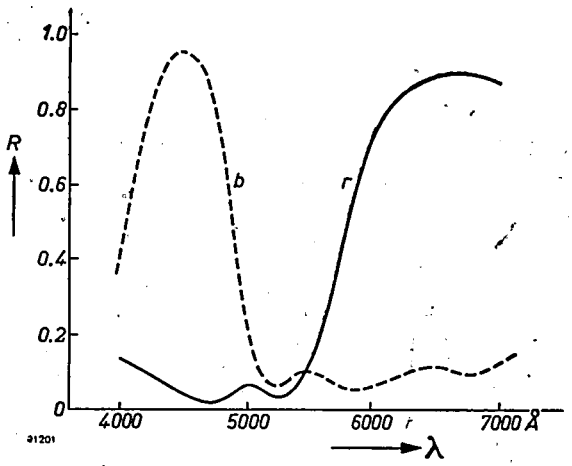


Fig. 6. Measured reflection coefficient R of two dichroic mirrors as function of the wavelength; r for a red, b for a blue reflecting mirror.

development of vacuum technique has stimulated sublimation methods, by which the material to be applied is produced in vapour form by cathode atomization or evaporation in high vacuum and deposited from the gas phase in solid form on the material to be coated. Particularly the evaporation method is much used. Not only metals, but also many salts can be applied in this way. This process is conditional on the material not decomposing below about 1500 °C and having a reasonable vapour pressure.

Our installation (*fig. 7*) is fitted up as follows. A dome K (*fig. 8*) rests on a table T and is sealed by a rubber ring R which is vacuum-tight when the dome is pumped out. Passing through the table around the pump tube a number of rigid molybdenum current leads I are mounted; these are electrically insulated from the table by a ceramic material and are capable of carrying 300 A. Evaporation elements V can be fixed to the conducting rods. The glass plate G to be coated rests on three rods S; two of these rods have a cap B at the top which can freely rotate and on the third rod this cap is fixed but the rod itself can be turned by a motor, for which purpose it passes through the table via a vacuum-tight bush L. When the motor is working, the glass plate rotates so that it can be coated very uniformly.

The evaporation elements V may consists of a crucible, heated by a tungsten spiral or a strip of tungsten or molybdenum. In many cases we use a type which is rather similar to an indirectly heated cathode. Hollows are turned in both top and bottom of a cylindrical piece of molybdenum (*fig. 9*). A tungsten heating coil, insulated with aluminium oxide is mounted in the lower cavity; the upper cavity, which serves as the crucible, contains the material to be evaporated. When the molybdenum

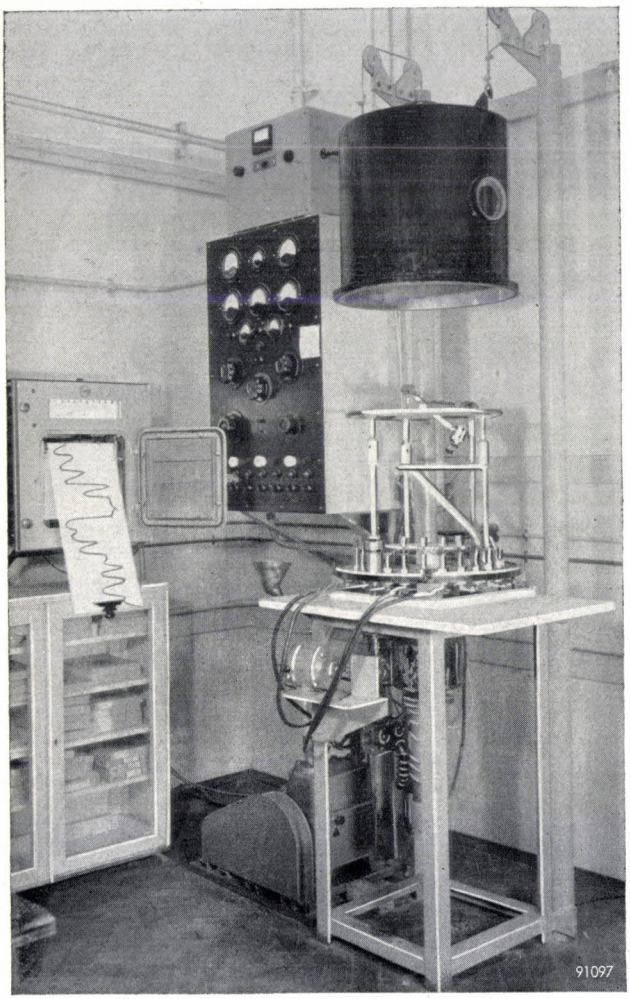


Fig. 7. Equipment for the sublimation of thin films in vacuum. The dome under which the sublimation is carried out (compare fig. 8) is shown raised; in the centre background is the switchboard for the control of the heating current to the evaporation elements. On the left can be seen a recording potentiometer which records the transmission of the glass plate under treatment as a function of the time.

is heated, the material evaporates and is deposited on the glass plate. In our apparatus twelve of these evaporation elements can be introduced and they can be heated independently. It is thus possible to deposit twelve successive layers without opening the dome. Also mounted under the dome are a lamp and a photo-cell by means of which the transmission of the plate can be measured during the sublimation process; part of a strip-chart recording of such a measurement can be seen in fig. 7.

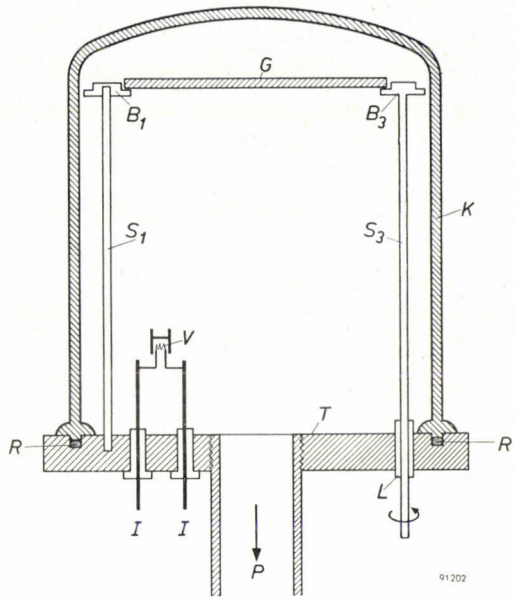


Fig. 8. Apparatus for sublimation of thin films. G is the glass plate being coated, K the dome, sealed vacuum-tight to the table T by the rubber ring R, P the exhaust tube, V one of the evaporation elements, with heater current leads I. S_1 fixed supporting rod with free cap B_1 (S_2, B_2 idem, not shown). S_3 rotating rod with fixed cap B_3, mounted in vacuum-tight seal L which leads it through the table; this rod is driven by a motor and causes plate G to rotate slowly.

The dome is suspended by steel cables passing over pulleys to counterweights. When air has been allowed to enter, the dome can thus easily be lifted up and the evaporation space is then accessible from every side.

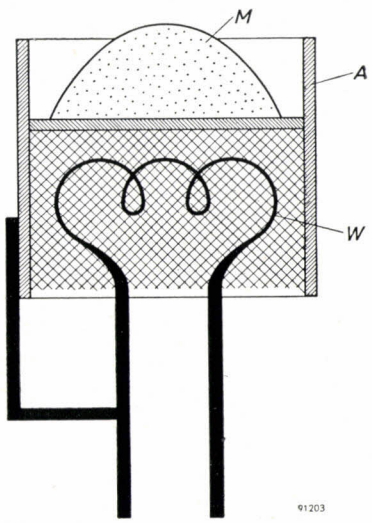


Fig. 9. Section of evaporation element. A molybdenum cup, W tungsten heater coil (insulated with aluminium oxide). M is the substance to be evaporated.

Applications of colour-separating mirrors

As will shortly be described in more detail in this Review, a colour-television picture can be obtained by superimposing three pictures, one in each of three basic colours (e.g. red, green and blue). One difficulty involved is that the three monochromatic pictures must be exactly equal in shape and size. Now if three projector tubes are placed beside one another in front of a screen (*fig. 10*), the middle one will throw a rectangular picture on the screen,

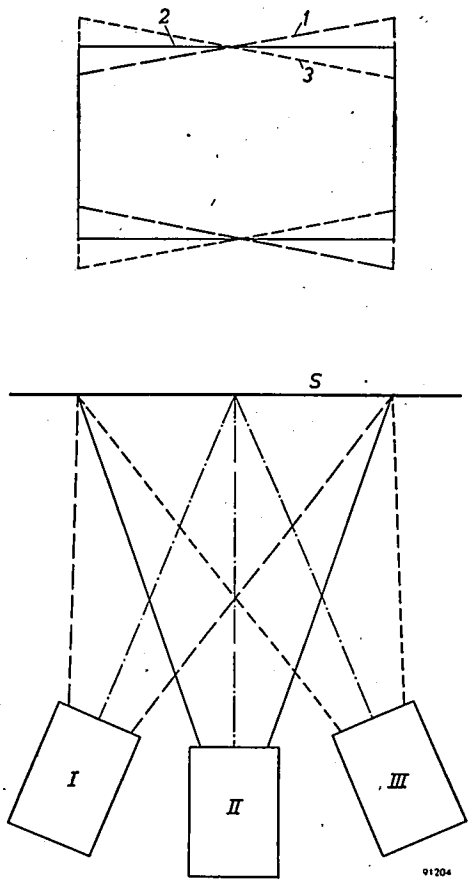


Fig. 10. The pictures from the three projectors *I*, *II* and *III* on the screen *S* do not properly coincide, since the projectors *I* and *III*, placed at an angle, give trapezium-shaped pictures (*1* and *3*), and the projector *II* a rectangular picture (*2*).

while the other two give trapezium-shaped pictures which deviate in opposite directions from the rectangle. There is no simple way in which to correct this fault by normal optical means in high aperture systems without losing a great deal in definition.

It is therefore preferable to project the pictures effectively from the same point. This can be achieved with the help of semi-reflecting mirrors (*fig. 11*). With normal semi-reflecting metal mirrors however, the loss of light would be very great. As already stated above, such mirrors reflect and transmit about 30%. About 30% of the light from projector *R* thus reaches the screen by reflection from mirror *r*, but only about 9% of the light from *G* and *B* is usefully employed. This lowers the intensity of the picture, so that bright colours are difficult to achieve. The result is better, but still unsatisfactory, when the mirrors are coated with SiO-TiO_2 (25% useful light instead of 9%).

Dichroic mirrors of the type discussed above, give a considerably more efficient use of the light. It can be seen from fig. 6 that the mirror *r* in fig. 11 reflects 88% of the red light from projector *R*, mirror *b* 93% of the blue light from projector *B*. Of this blue light only 6% is lost by reflection from mirror *r*, so that 87% reaches the screen. The green light from projector *G* loses 10% at mirror *b* and 10% at mirror *r* by reflection, so that 81% is still effective.

In the arrangement in fig. 11, the mirror *r* is much larger than mirror *b*. This is necessary because *r* is situated much further from the projectors *G* and *B* and the beam therefore has a larger cross-section at *r* than at *b*. In order to make *r* smaller, preference is often given to an arrangement by which the mirrors cut one another, forming a cross (*fig. 12*). Here use is made of the fact that not only does blue light suffer little loss in passing through a red-reflecting mirror, but also that red will pass through a blue-reflecting mirror without much loss (see fig. 6). The green light in fig. 12, as in fig. 11, passes through both mirrors.

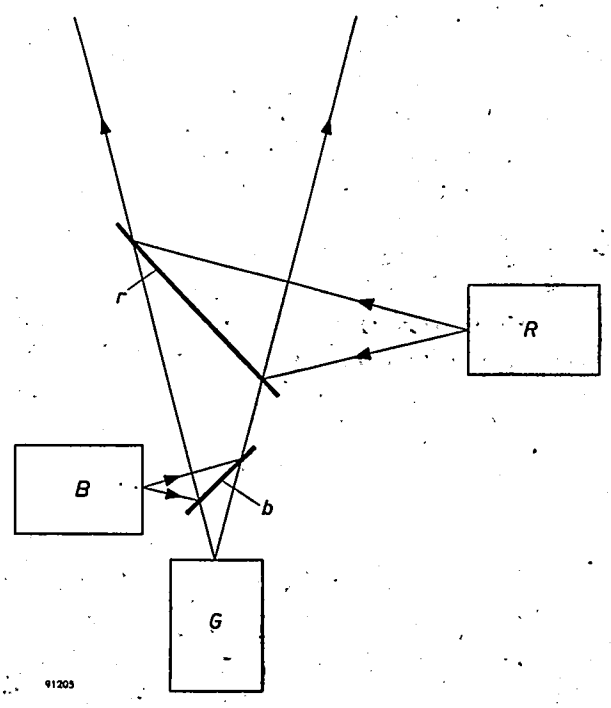


Fig. 11. With the use of semi-reflecting, dichroic mirrors, the beams from the three projectors (*R* red, *G* green, *B* blue) can be made to coincide, with little loss of light. *r* is a red, *b* a blue reflecting mirror.

In the last example, light from three monochromatic sources was combined to give one coloured picture. The reverse can also be done, that is, the light from a multi-coloured object can be split into three basic colours by means of a similar combination

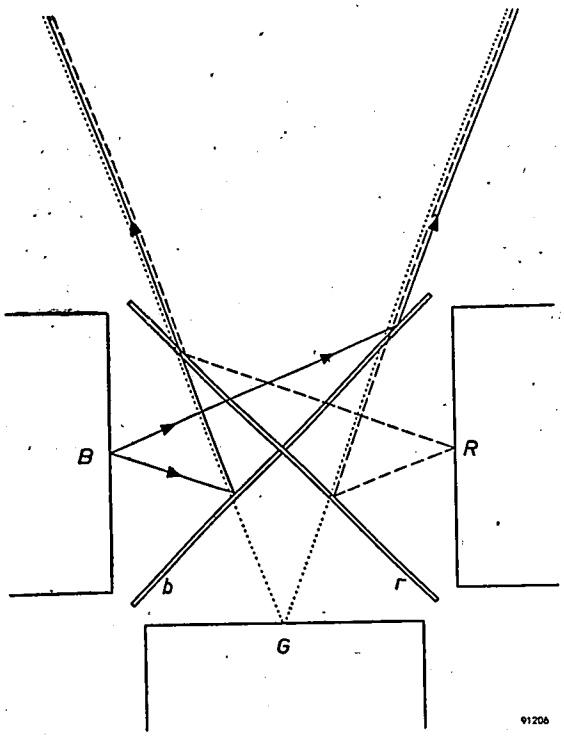


Fig. 12. As fig. 11, but with the arrangement modified so that the red reflecting mirror *r* need be no larger than the blue reflecting mirror *b*.

of mirrors. This is done in some colour television cameras. Apart from the more efficient use of the available light, another important point is that the scene is viewed from one place, not three, and in all colours at the same instant.

Corrective measures

Since the mirror must be accurately plane, the glass plate which carries the interference layers must be fairly thick. This thickness makes certain corrections necessary.

In the first place, the thickness prevents light reflected from the back surface from coinciding with that reflected from the front surface, so that double images occur. The image originating from the back surface is of the complementary colour and is not very weak, since it is of the colour which is transmitted strongly by the interference layer. This disturbing reflection can be weakened by providing the back of the mirror with a film with low reflection properties for the colour of the transmitted light.

In the second place, the thickness of a plane-parallel glass plate placed at an angle in a non-parallel (thus convergent or divergent) beam of light, gives rise to optical aberrations. In *fig. 13* we see how ray *1* is but slightly displaced sideways by the plate while ray *3* is displaced to a much greater degree. This causes asymmetric focusing at the image point *B*. Rays obtained by rotating ray *1* about the normal *LM* (*M* is the point where the incident rays, if produced, intersect), all converge to the point *N* of intersection between *LM* and the chosen ray *1*. This intersection lies closer to the plate than *B*. In this way mainly astigmatism and slight coma occur [2]). These faults can be compensated by introducing a separate optical element, but that would mean that, for instance, a cylindrical lens would have to be placed in the path of the rays to correct the astigmatism. H. de Lang, of the Philips Laboratory at Eindhoven, has shown [3]) that a compensation can be achieved merely by altering the form of the glass plate slightly. The front surface must remain plane because of the mirror action, but if the plate is made slightly wedge-shaped or if the back surface is ground weakly convex, the aberrations can be suppressed to a great extent. For a television camera with a plate a few mm thick at an angle of 45° to the light beam, a prism angle of 7′ or a radius of curvature of 8 m for the back surface is sufficient.

In the third place, the thickness of the glass introduces difficulties in the crossed arrangement of the mirrors. If one uses the system shown in fig. 12, the mirror *b* must be interrupted to allow mirror *r* to pass through and therefore no blue

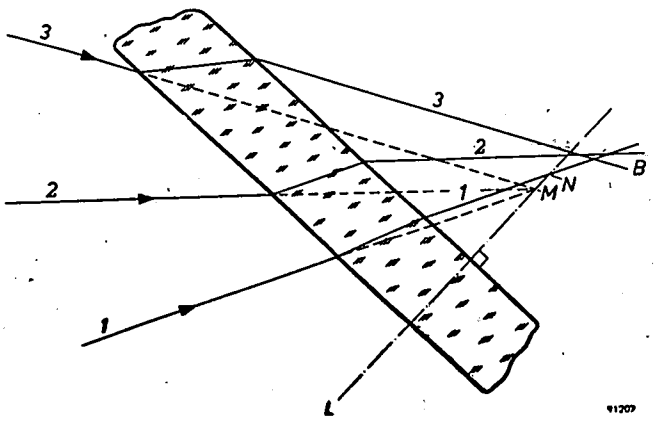


Fig. 13. The thickness of the glass of the semi-reflecting mirror produces aberrations in the transmitted beam: the rays *1*, *2* and *3* converging towards *M* no longer intersect in one point after passing through the mirror.

[2]) For these optical aberrations see e.g. W. de Groot, Philips tech. Rev. **9**, 301-308, 1947/48.

[3]) See Philips Res. Rep. **12**, 181-189, 1957 (No. 3).

light would be reflected where the cross-over occurs. A construction as shown in *fig. 14a* is thus usually preferred. The thickness of the glass means that the boundary *AB* forms an irregularity which can intercept light, as in the case of ray *1*. Such a cross is therefore usually placed in a position in the apparatus where the rays travel as far as possible parallel to the plane *AB*, e.g. parallel to the optical axis: ray *2* in fig. 14*a* is not intercepted, but the rays which do not travel parallel to the axis (originating for example from the top and bottom edges of the object to be reproduced), like ray *1*, may be partly intercepted and shadow-formation by plane *AB* may occur. This can be prevented by making the first mirrors (r_1, b_1) thicker than those behind (r_2, b_2; fig. 14*b*): refraction in the thicker mirrors displaces the beam so far laterally that no light falls on the dividing surface between the back mirrors. This method, however, is only permissible for objects which are apparently at an infinite distance, because otherwise each image point would be doubled.

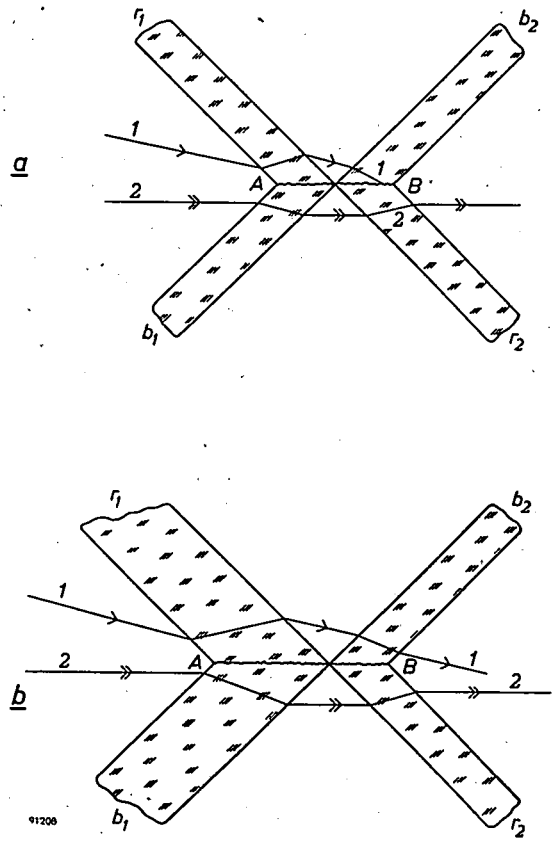


Fig. 14. $r_1 r_2$-$b_1 b_2$ are dichroic mirrors in an arrangement differing slightly from that in fig. 12. In (*a*) ray *2* which enters parallel to the boundary plane *AB*, can pass unhindered, but the obliquely incident ray *1* is lost in *AB*. In (*b*) the first mirrors (r_1 and b_1) are made so thick that ray *1* is laterally displaced far enough to avoid *AB*.

Summary. The well-known phenomenon of the interference of light in thin films has found many practical applications in recent years. According as the refractive index of the film is higher or lower than that of the carrier substance, the film increases or reduces the reflection. The magnitude of the effect depends on the relationship between the thickness of the film and the wavelength of the light used. By a suitable combination of layers, the effect can be enhanced and in this way practically loss-free mirrors can be prepared which transmit one region of the visible spectrum and reflect another region (dichroic mirrors). The preparation of such layers by sublimation in vacuum and the application of dichroic mirrors in colour television are discussed.

A HIGH-PRECISION LATHE HEADSTOCK

621.941.26-229.331

Lathes for machining with high precision in a direction perpendicular to the axis of rotation are generally available. On special toolroom lathes shafts may be turned to a diameter tolerance of a few microns, whilst local surface roughness may even be reduced to e.g. $\frac{1}{2}$ μ.

Whenever a lathe is used to machine flats or profiles on end surfaces, e.g. plastic lenses, we are concerned with the accuracy in a direction *parallel* to the axis. This accuracy, which is not very critical for ordinary applications, is mainly dependent upon the thrust bearings of the lathe spindle.

With other machine tools, such as grinding and drilling machines, it may also occur that for obtaining extremely smooth surfaces a highly accurate thrust bearing is desired. Adequate results for this purpose may be reached with conical slide-bearings, but a fairly large tolerance must be observed to allow for changing loads. Clearly this affects the accuracy that can be attained. A bearing free of this drawback and ensuring a very high degree of axial and radial accuracy has been developed in the Eindhoven Research Laboratories. Its construction is shown in *fig. 1*. The end of the main spindle, which is accurately supported in radial direction by the bearing, is fitted with a piece of "improved wood" (B). This wooden member rests against a stationary metal thrust block, whose front surface has been lapped plane and perpendicular to the main spindle with a high degree of accuracy. The main spindle is pressed against the thrust block with a *constant force*, owing to the fact that a constant pressure below atmospheric pressure is maintained in the closed compartment D in which this contact takes place. The thrust block can be axially displaced by means of a micrometer screw. It this way the workpiece on the main spindle can be displaced over an extremely small distance (the cutting depth) with respect to the tool.

Evidently the position of the tool in axial direction must be fixed with great precision as well, but this requirement can be readily satisfied.

The low pressure in the above compartment also sucks in an ample flow of oil along the main spindle, so that the latter is efficiently lubricated, which is necessary for a very accurate radial support: both the steel spindle and the three supporting surfaces (F in fig. 1) of the bronze bearing block have been accurately lapped to size with a tolerance of only about 1 μ. This means that a temperature drop of 5 °C would cause the spindle to seize because of the difference in coefficient of expansion between the two materials. This is avoided by keeping the whole bearing at a constant temperature of about 30 °C with a heating element. The enforced oil-flow prevents any excessive local overheating of the sliding surfaces, which would increase the play.

The spindle should be driven with the greatest uniformity, avoiding any vibrations in the drive.

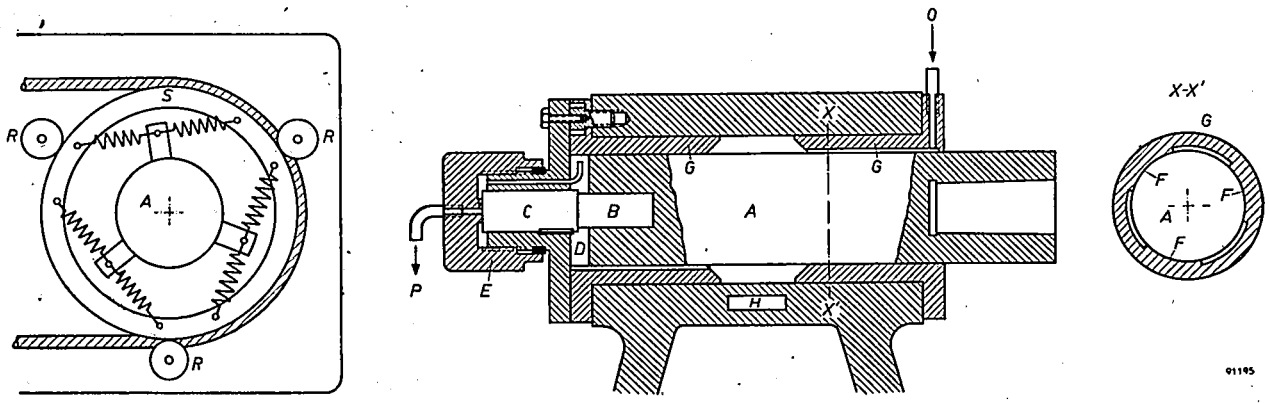


Fig. 1. Construction of the precision headstock. A steel shaft running in bronze bearing block G (which has three bearing surfaces F, see section XX' on right). B improved-wood end-piece; C metal thrust bearing block, which may be moved axially by means of the micrometer screw E. In the space D a constant pressure, beneath atmospheric, is maintained. H heating element. O oil inlet, P oil outlet.

The left-hand figure shows the drive. The pulley S, connected to the headstock shaft by six springs, runs in three rollers R mounted on a bracket attached to the wall of the workshop.

The pulley is therefore not fitted rigidly to the spindle, but coupled to it via steel springs, the pulley being supported not by the spindle bearing, but on rollers of its own.

The axial accuracy attainable with this construction can be deduced from *fig. 2*, a photo-micrograph

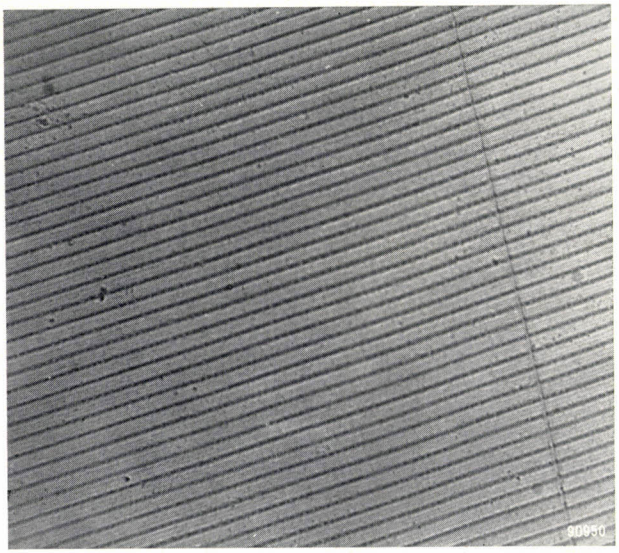


Fig. 2. Photomicrograph ($\sim 250\times$) of the surface of a disc of plastic material machined flat with a diamond tool on a precision lathe provided with the headstock described here. From the uniformity of the groove width it can be deduced that movement in the axial direction during working was less than 1/200 micron.

of the surface of a flat disc machined on a lathe equipped with the headstock in question. A diamond-tipped tool was used with a round tip of radius 200 μ. The lateral displacement of the tool was 9 μ per revolution. The finished surface thus consists of adjacent grooves 9 μ wide and with a depth of $4.5^2/(2\times 200) \approx 1/20$ μ. A close examination of the photo shows that the width of the groove never varies by more than about 5%. This means that during a short period (the machining of the area shown in the picture lasted about 5 seconds) the axial displacement of the lathe spindle relative to the tool did not amount to more than 10% of the groove depth, i.e. less than 1/200 μ.

The same lathe was used for machining a cavity resonator for 8.75 cm wavelength, earlier described in this Review [1]), for high-precision measurements of the dielectric losses in solids. It was essential to keep local unevenness of wall, bottom and cover of the cavity as small as possible with respect to the penetration depth of the H.F. currents, viz. about 1 μ for the wavelength in question. Measurements of the quality factor of the cavity resonator confirmed that the unevenness of the machined surfaces was indeed very small with respect to the penetration depth.

This lathe is now also in use for making the aspherical Schmidt correction plates of mirror cameras for X-ray fluorography [2]).

L. M. LEBLANS.

[1]) M. Gevers, Measuring the dielectric constant and the loss angle of solids at 3000 Mc/s, Philips tech. Rev. **13**, 61-70, 1951/52.

[2]) Philips tech. Rev. **13**, 269-281, 1951/52. For manufacturing the correction plates of smaller Schmidt systems (e.g. for projection TV sets), a simpler method has been developed; see Philips tech. Rev. **9**, 349-356, 1947/48.

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.

2432: J. de Jonge and H. J. van den Bergh: The preparation of 2,4,6-trimethoxymethyl phenol (Rec. Trav. chim. Pays-Bas **75**, 1214, 1956, No. 9-10).

The preparation of 2,4,6-trimethoxymethyl phenol is described.

2433: P. Jongenburger: De invloed van dislocaties op enige fysische eigenschappen van vaste stoffen (Ned. T. Natuurk. **22**, 297-305, 1956, No. 10). (The effect of dislocations on certain physical properties of solids; in Dutch.)

A review of the effect of dislocations in metals (especially copper) on the electrical resistance and on the increase of resistance in a magnetic field. The effect on nuclear spin resonance is also considered; this is an effect which can be studied also in non-conducting crystals. Measurements of these effects enable the concentration of dislocations to be approximately determined.

2434: J. I. de Jong, R. Dijkstra and J. de Jonge: The ortho/para distribution of methylol groups in the reaction of phenol and formaldehyde, I (Rec. Trav. chim. Pays-Bas **75**, 1289-1302, 1956, No. 9-10).

In the reaction between phenol and formaldehyde in aqueous solution at 90 °C and $p\mathrm{H} = 5.5$-6.5,

about 58% ortho substitution and 42% para substitution takes place initially. A colorimetric method for the determination of methylol phenols in solution is described.

2435: D. Kleis: Toepassing en opstelling van luidsprekers in geluidsinstallaties (T. Ned. Radiogenootschap 21, 237-254, 1956, No. 5). (Application and arrangement of loudspeakers in sound installations; in Dutch.)

Speech and music put different requirements on the acoustics of an auditorium, so neither for live performances nor for recordings can the listening conditions be optimum for both. Inadequate acoustical conditions can be corrected both for live performances and recordings by using separate loudspeakers for direct and indirect sound. Loudspeakers for direct sound should have pronounced directivity towards the listeners so as to minimise reverberation. Sound reinforcement or sound transmission by this type of loudspeakers — horns or sound-columns — yields optimum intelligibility of speech; a stereophonic reproduction can fully represent the direct sound of a live performance of speech or music. Loudspeakers around the auditorium, producing diffuse reverberating sound are used to suggest reflections from the walls and to produce the adequate level of indirect sound for music. In order to suggest natural reflections, their signals have to be retarded in accordance with the loudspeaker positions. The acoustics of the auditorium may be controlled by the studio and so be made optimum for any kind of performance by using two-channel transmission or twin-track recordings.

2436: J. H. Stuy: Studies on the radiation inactivation of micro-organisms, IV. Photoreactivation of the intermediate stages in the transformation of *Bacillus cereus* spores into vegetative cells (Antonie van Leeuwenhoek J. Microbiol. Serol. 22, 337-349, 1956, No. 4).

Continuation from articles I, II & III (these abstracts 2274, 2428, 2429). The transformation of *Bacillus cereus* spores into vegetative cells has been studied. It was possible to distinguish between five different stages. Each stage could be isolated. During the development of stage III cells from stage II spores the phenomenon of photoreactivation could be observed. The results strongly suggest that the non-photoreactivability of UV-inactivated resting spores and stage II spores is due to their limited metabolic activity.

2437: M. G. van der Steeg and K. J. de Vos: Reversibility of the coercive force in Alnico 5 (J. appl. Phys. 27, 1250, 1956, No. 10).

Interim report on the discovery by means of X-rays of a new face-centered cubic phase in Fe-Co-Ni-Al magnetic alloys. The new phase is observed after long annealing (70 days) of the alloy at 600 °C. As is well known, a growing phase exerts an influence on certain physical properties before its presence can be demonstrated by X-ray diffraction. Magnetic measurements made after much shorter annealing times suggest that the newly discovered phase plays an important part in the achievement of the large coercivity of Alnico 5 (Ticonal G).

2438: L. J. van der Pauw and D. Polder: The photo-thermoelectric phenomenon in semiconductors (J. Electronics 2, 239-240, 1956, No. 3).

The photo-thermoelectric effect is the phenomenon that the thermoelectric potential between two materials changes when one of the materials is illuminated. For this effect an expression has been derived by Tauc, assuming that the energy difference between the conduction band and the valency band is independent of the temperature. The authors show, however, that Tauc's result is valid only if the mobility of holes and electrons are also independent of the temperature. A new expression is given which takes both temperature dependencies into account. The additional terms in this expression are of the same order of magnitude as Tauc's result.

2439: H. G. van Bueren: Magnetoresistivity of plastically deformed metals (Proceedings of the IUTAM Colloquium on the deformation and flow of solids, Madrid 1955, published by Springer, Berlin 1956, pp. 79-89).

The magnetoresistance at very low temperatures of pure polycrystalline copper has been found to increase considerably, when plotted in a reduced diagram, upon plastic deformation. This effect is due to dislocations, as subsequent annealing up to 220 °C has no effect whatever on the additional relative magnetoresistivity. Neither vacancies and interstitials nor impurity atoms effect the relative magnetoresistance, and the latter quantity forms therefore a suitable tool to study dislocations separately. A theory is outlined, according to which the anisotropic scattering by dislocations is the main cause of the phenomenon. The numerical agreement between the result of this theory and the observations is satisfactory.

2440: F. A. Kröger and H. J. Vink: Relations between the concentrations of imperfections in crystalline solids (Solid State Physics, **3**, editors F. Seitz and D. Turnbull, pp. 307-435, Academic Press, New York 1956).

A number of properties of crystalline solids such as the electronic or ionic conductivity, the colour, the luminescence, the magnetic susceptibility, are determined by the presence of imperfections, viz. free electrons and holes, vacant lattice sides, interstitial atoms, and foreign atoms. The concentrations of these imperfections are not independent of each other. In the present paper a quantitative discussion is given of the relations existing between the various imperfections. Mainly complete equilibrium (corresponding to high temperature) is discussed. Starting from this state, possibilities for incomplete equilibrium (as may be present after cooling) are also considered. The discussion covers *a*) pure compounds having a composition corresponding to a simple stoichiometric ratio, *b*) compounds showing a deviation from the simple stoichiometric ratio, *c*) compounds containing one or more types of foreign atoms. With the aid of the theory given — which is based on the theory of imperfections of Schottky and Wagner — it is possible to give a quantitative interpretation of the intricate relation existing between the conditions of preparation of crystals, and their electrical, optical and magnetic properties. A special aspect of these relations is the interdependence of the solubilities of foreign atoms.

2441: B. D. H. Tellegen: La recherche pour une série complète d'éléments de circuit idéaux non-linéaires (Rendiconti del seminario matematico e fisico di Milano **25**, 134-144).

The five ordinary circuit elements (resistance, inductance, capacitance, transformer, gyrator) can be considered, by definition, to represent the simplest linear constant passive systems. From a physical point of view, linear constant active systems, linear variable systems, and non-linear systems all belong to the class of non-linear constant passive systems. The search for the simplest non-linear constant passive systems leads to ideal non-linear net-work elements. For order zero these are the ideal rectifier, the elements called voltage sink and current sink, and four ideal amplifiers.

2442: H. G. van Bueren: Dislocaties in niet-metalen (Ned. T. Natuurk. **22**, 394-402, 1956, No. 12). (Dislocations in non-metals; in Dutch.)

Dislocations have much more influence on the physical properties of alkali-halides and semiconducting elements than on those of metals. They can act as sources of vacancies, and therefore of colour centres, in ionic crystals, which explains the enhanced colouring by X-ray irradiation of previously cold-worked crystals, and the equally enhanced conductivity. In germanium, dislocations produce acceptor levels and reduce the carrier mobility. The motion of dislocations in valence type crystals, is impeded by the typical core structure of the imperfections, which is characterized by the occurrence of unpaired bond electrons.

2443: J. Bloem, A. J. van der Houven van Oordt and F. A. Kröger: A new luminescence emission in Cu_2O (Physica **22**, 1254-1256, 1956, No. 12).

Cu_2O is known to show luminescence in the infra red near 1 μ. In the present investigation it is shown that at 20 °K this emission consists of two sub-bands at 0.91 and 1.01 μ. Samples prepared at 960 °C under oxygen pressures of 10^{-1}-10^{-3} mm Hg show in addition two luminescence peaks at shorter wavelengths, viz. 0.72 μ and 0.82 μ. The former bands are attributed to copper vacancies, the latter to oxygen vacancies.

2444: J. A. Kok and M. M. G. Corbey: Breakdown of liquid insulating and dielectric material (Appl. sci. Res. **B6**, 197-206, 1956, No. 3).

One of the possible causes of long-time breakdown of low-frequency impregnated paper capacitors and cables is known to be the long-time formation of a conducting bridge between the electrodes, followed by a short-time thermal breakdown. Such a bridge may consist of polarizable particles, ions or colloidal contaminants of high dielectric constant, particularly conducting material. These particles flow towards a place of maximum stress because of the action of transverse quadratic gradient forces in a non-uniform field, thus forming a conducting path for the short-time development of gas and the subsequent breakdown by ionization. From a long-time breakdown strength of 50 kV/mm there follows a particle diameter of about 30 Å.

2445: J. A. Kok: Theory and probe measurements of Gabor's gas-filled triode (Appl. sci. Res. **B6**, 207-221, 1956, No. 3).

By means of a number of probe measurements in Gàbor's gas-filled triode with a spread dark discharge in the cathode chamber, the form of the characteristic I_a-V_a curves was established and

several relations between various quantities derived. With these results it was possible to calculate a relation between Gàbor's and Fetz's criteria of the transition of the spread discharge into a constricted one.

2446: J. Ugelstad and J. de Jonge: The alkaline hydrolysis of alkoxymethyl ureas (Acta chemica scandinavica **10**, 1475-1487, 1956, No. 9).

Alkoxy methyl ureas, surprisingly enough, were found to be quite readily hydrolysed in moderately alkaline solution at room temperature, giving urea formaldehyde and an alcohol as end products. The decomposition of different alkoxymethyl ureas was investigated kinetically. The hydrolysis was found to be proportional to the OH-ion concentration. Based on kinetic investigations in strongly alkaline solution, a choice is made between two possible reaction paths for the reaction. It may be concluded that the reaction involves the breaking of an other bond, i.e. that methylol urea and alcohol are the first decomposition products. The reaction is specific for ethers of methylol urea, ethers of methylol compounds of carboxylic acid amides being comparatively very stable in alkaline solution. The reaction mechanism is discussed.

R 314: H. de Lang: Compensation of aberrations caused by oblique plane-parallel plates (Philips Res. Rep. **12**, 181-189, 1957, No. 3).

Methods are given for the correction of aberrations (astigmatism and coma) introduced by an oblique plane-parallel plate in a convergent beam. It is shown that astigmatism can be eliminated, either by giving the plate a slight amount of prismaticy or by giving the rear surface of the plate a slight amount of spherical or cylindrical curvature. In practical cases the first method is preferable as it avoids cylindrical surfaces and also nearly eliminates coma. Off-axis correction of astigmatism impose the extra condition that the derivative of the astigmatism in the meridional section be zero. It is shown that in general this on-axis correction can be achieved by a suitable combination of prismaticy and curvature.

R 315: H. G. van Bueren: Influence of lattice defects on the electrical properties of cold-worked metals (Philips Res. Rep. **12**, 190-239, 1957, No. 3).

Continuation of **R 306.**

R 316: W. Ch. van Geel and B. J. J. Schelen: Some properties of oxide layers produced on aluminium by electrolytic oxidation (Philips Res. Rep. **12**, 240-248, 1957, No. 3).

A description is given of measurements of the thickness (d), dielectric constant (ε) and density (ϱ) of oxide layers produced on Al by electrolytic oxidation. The X-ray investigation of the structure of these layers was also undertaken. Two different values were found for the thickness of the layer, viz. $d = 13.7$ Å/volt and $d = 12.7$ Å/volt. The former value was found after chemical cleaning, the latter without any prior removal of the oxide layer formed by reaction with the air. The cleaning probably causes an increase of about 7% in the surface area. If this is correct, then $d = 12.7$ Å/volt is the more acceptable value. For ϱ we found the value 3.1 g/cm^3, for ε the value 8.7. When it is formed at room temperature, the oxide is completely amorphous, except when a layer of boehmite has been produced on the Al by boiling it in water. In the latter case γ'-Al_2O_3 is also formed. Oxidation at 100 °C always gives some γ'-Al_2O_3 as well as amorphous Al_2O_3.

R 317: J. L. H. Jonker: On the theory of secondary emission of metals (Philips Res. Rep. **12**, 249-258, 1957, No. 3).

The angular distribution of the secondary emission of nickel is measured with small intervals of energy. The results allow us to make a hypothesis about the mechanism of the emission, about the energy distribution, and about the scattering of secondary electrons inside the metal. An explanation is proposed for the cosine distribution of the emitted electrons.

R 318: K. F. Niessen: On a higher approximation of the critical field strength for an antiferromagnetic (Philips Res. Rep. **12**, 259-269, 1957, No. 3).

The critical magnetic field strength is derived for an antiferromagnetic with different anisotropy constants for the two sublattices, account being taken of the field-strength dependence of the parallel and perpendicular susceptibilities and a higher approximation of the anisotropy energy. Special attention is paid to the variation of the critical field strength for a small deviation of the external field from the preferential axis. Formulae (20) and (21) contain the results for the case of equal and (25) and (27) for the case of unequal anisotropy constants.

N.V. PHILIPS' GLOEILAMPENFABRIEKEN

VOLUME 19, 1957/58, No. 3 pp. 73-108 Published 9th October 1957

Philips Technical Review

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

"FRENA", A SYSTEM OF SPEECH TRANSMISSION AT HIGH NOISE LEVELS

by F. de JAGER and J. A. GREEFKES.

621.395.665.1

Much of the development in telecommunications is concerned with improving the quality of the link, for example by employing frequency modulation or pulse-code modulation systems. In mobile installations a different aim is often pursued: sometimes a certain minimum of quality, just high enough for the intelligible transmission of speech, is accepted, but it is desired that the signals shall continue to be intelligible at very high noise levels, i.e. with very low transmitting power or over a very great distance, or both. For this minimum quality to be obtainable, the noise power (with single-sideband modulation) must not normally exceed 10% of the signal power received. In two variants of the "Frena" experimental system, however, the permissible limit has been raised to 16% and 25%, and in the variant known as "Frenac" it has been raised to as high as 40% of the signal power.

Introduction

The continuing development of telecommunications is largely a struggle with the inevitable noise and interference. These unwanted voltages, consisting mainly of crosstalk and circuit noise in telephony and of atmospherics, aerial noise and transmitter interference in radio, television and radar, make reception impossible beyond certain distances. If the signal carrying the information becomes so weak that it is swamped by interference, amplification at the receiving end has no value, since the interfering voltage is amplified along with the signal. Matters are even worse in actual practice because the amplifier itself contributes to the noise, mainly from the first amplifying tube.

In radio links and in radar, attempts can be made to reduce to a minimum the interference coming from the medium of transmission, by using directional aerials, for example, and in line telephony a number of measures can be adopted to suppress crosstalk. But ultimately a certain minimum interference level is reached which cannot be lowered any further. A similar state of affairs applies to the noise contributed by the receiver. It has been possible greatly to reduce the latter by designing tubes with good noise properties, and this line of development has by no means come to an end, particularly in the design of tubes for very high frequencies. A lower level is also arrived at for noise arising in the receiver (related, *inter alia*, to the properties of the cathode). For a given minimum level of interference the distance over which transmission is possible (extreme range) can be determined from two factors. These are the *maximum transmitting power*, as imposed by economic or other considerations, and the *minimum signal-to-noise ratio* at which reception is still practicable.

The latter point — the minimum admissible signal-to-noise ratio — has been the subject of a number of developments which have led to increases in the ranges over which communication is possible. Two lines of development may be specially mentioned. On the one hand systems have been developed whereby the interference in the frequency band of the original signal is to some extent reduced, and the signal-to-noise ratio at the output consequently improved by employing a considerably greater bandwidth for the transmission path. In these systems transmission remains linear, i.e. the signal undergoes no fundamental distortion, and any kind of information can be transmitted, including music. On the other hand transmission systems have been developed for special kinds of information such as telegraphic signals or speech; in these systems, the specialized character of the signals makes it possible

to reduce interference without greatly increasing the bandwidth of the transmission path. In the latter class we may include the system of signal correction which has long been employed in Morse telegraphy and which re-appears in a refined form in the T.O.R. system for teleprinter links [1]). We may also include the practice of peak-clipping in the speech signal, which has long been usual in certain mobile radiotelephony installations (it can be regarded as a method of using available transmitter power in a more effective manner). To this same class belong the "Frena" system and its variants, which have been developed in recent years in the Eindhoven laboratories of Philips and which we shall describe in the present article. This system is also mainly intended for speech transmission via mobile radio installations (not for normal line telephony); here linearity of transmission is abandoned for the sake of suppressing interference.

To get a good grasp of the subject, it will be desirable first to treat briefly the linear systems; in particular we shall discuss the degree of noise suppression obtainable from them. (In what follows we shall refer only to interference in the form of *noise*, since this lends itself best to quantitative study.)

The suppression of noise in various linear transmission systems

The noise-suppressing systems we would include in this class are, first and foremost, frequency modulation and pulse-position modulation, and secondly the systems known as pulse-code modulation and delta modulation, in which the information is carried by a series of pulses having magnitudes of zero or unity. The properties of the coded systems are particularly good for communication via a large number of relay links.

The common feature of all these systems might be described as follows. In any transmission path the total noise power is proportional to the bandwidth. For a given transmission path (this is always considered to include the receiver's first stage, which contributes to the noise) and for a given transmitting power, the signal received under conditions of optimum matching has an "intrinsic" signal-to-noise ratio; this is the received signal power divided by the power of the noise from the transmission path *in a frequency band having the width of the audio signal*. In the systems in question, the audio signal is converted at the transmitter into a signal having a much wider frequency band, by a suitable modulation method. At the receiving end, a suitable detection method recovers the original narrow-band signal from the incoming wide-band signal, only a fraction of the noise from the wide band being detected along with the signal; the result is that the signal-to-noise ratio at the output, which we might term the "effective" ratio, is higher than the intrinsic one.

The systems named above differ considerably in the extent to which they allow of improvements in the ratio of signal power to noise power consequent on increases in the bandwidth of the transmission path. In frequency modulation and pulse-position modulation the improvement increases as the square of the bandwidth [2]); in delta modulation it increases approximately with the fifth power of the bandwidth [3]); and in pulse-code modulation the increase is exponential [4]). Information theory shows that the last-mentioned is the steepest that can be attained.

The improvements indicated above only apply so long as the noise level in the transmission path (inclusive of the first receiver tube, as before) does not become too high. With increasing noise, a point is reached in all these systems at which a further increase will interfere with their particular method of detection to such an extent that the signal will loose its coherency almost completely. In frequency modulation that point is certainly reached when the noise is so strong that the polarity of the signals at the input of the discriminator is often reversed. The point is reached in the pulse systems when additional pulses originating in noise begin to occur. In this connection it is possible in all these systems, given a certain noise level, to define an optimum bandwidth for the transmission path. Where bandwidth is greater than optimum, the higher noise voltage thus fed to the input of the detector destroys the coherency of the received signal; with a bandwidth smaller than optimum, on the other hand, full advantage is not taken of the noise-suppressing properties of the system.

In any system, therefore, and for a given intrinsic signal-to-noise ratio $(S/N)_i$, there exists a certain maximum value for S/N, the effective signal-to-noise ratio at the output. The maximum value can only be obtained by making use of the optimum bandwidth: see *fig. 1*, which refers to frequency

[1]) H. C. A. van Duuren, Typendruktelegrafie over radioverbindingen (T.O.R.), T. Ned. Radiogenootschap **16**, 53-67, 1951 (in Dutch). The basis of this system is that code-letters are automatically repeated whenever they have not been properly received.

[2]) S. Goldman, Frequency analysis, modulation and noise, McGraw-Hill, New York 1948.

[3]) F. de Jager, Delta modulation, a method of p.c.m. transmission using the 1-unit code, Philips Res. Rep. **7**, 442-466, 1952.

[4]) S. Goldman, Information theory, Prentice-Hall, New York 1953.

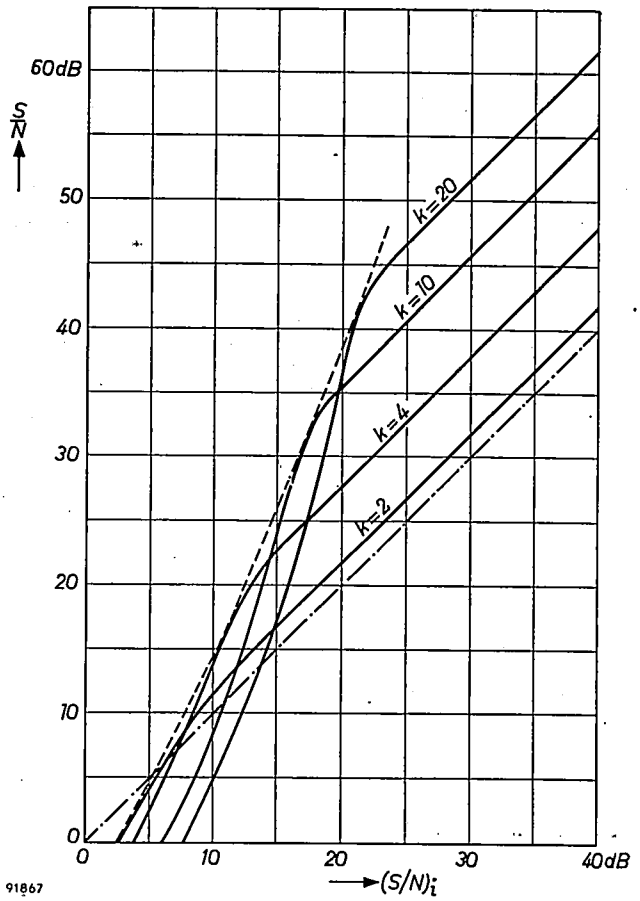


Fig. 1. S/N, the effective signal-to-noise ratio obtained with frequency modulation, as a function of $(S/N)_i$, the "intrinsic" signal-to-noise ratio (based on data given by F. L. H. M. Stumpers, Proc. Inst. Rad. Engrs. **36**, 1081-1092, 1948). The frequency band of the transmitted signal has k times the width of that of the original signal. The curves for different k values intersect; thus an envelope (the broken-line curve) can be drawn indicating the limit of what can be accomplished with frequency modulation. The chain-dotted curve refers to single-sideband modulation, which is usual in carrier telephony and which gives an effective signal-to-noise ratio equal to the intrinsic ratio.

modulation. The four systems named above have been subjected to a thorough analysis, particularly in respect of high noise levels. Curves were drawn expressing the theoretical maximum signal-to-noise ratios obtainable in direct links as a function of the intrinsic signal-to-noise ratio. It was found that the curves for the four systems were closely bunched over a wide range [5]). To put it another way, the noise level at which the system starts to break down (or ceases to satisfy certain demands made on it) is about the same for all the systems.

This situation is illustrated in *fig. 2*. In each of the four systems the relationship between S/N and $(S/N)_i$ depends on the bandwidth that is chosen, but, on account of the optimum condition mentioned above, the values appropriate to the performance that can actually be obtained from the system always lie to the right of the curve in fig. 2. If, for example, an effective signal-to-noise of 30 dB be required, then an $(S/N)_i$ of at least 17 dB will have to be available; in other words, reception is only possible over a distance such that the power of the received signal is fifty times as great as the noise power within the bandwidth of the low-frequency signal. For the sake of comparison the relationship between S/N and $(S/N)_i$ in the case of straightforward single-sideband modulation, which is normally employed in carrier telephony [6]), is also drawn in fig. 2 (chain-dotted line). Here the effective signal-to-

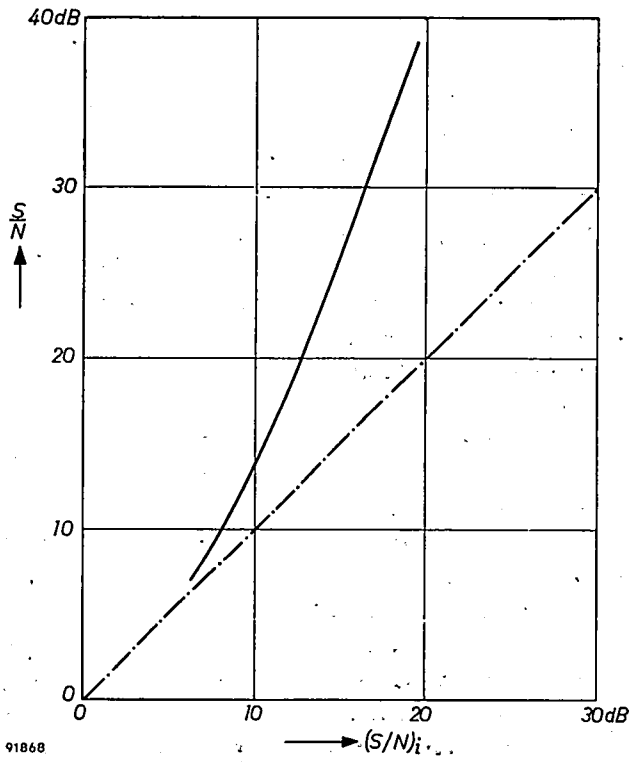


Fig. 2. The maximum value of S/N, the effective signal-to-noise ratio at the output, obtainable from one of the four linear interference-suppressing systems (frequency modulation, pulse-position modulation, pulse-code modulation and delta modulation), as a function of $(S/N)_i$, the intrinsic signal-to-noise ratio. The curve is the median of four envelopes drawn for the four systems in the same manner as in fig. 1 [5]).

The chain-dotted curve relates to single-sideband modulation. The curve corresponding to double-sideband modulation, as employed in broadcasting, lies considerably lower; at $(S/N)_i = 6$ dB, for example, S/N is −1 dB, so that in this case the maximum signal power at the output is lower than the average noise power.

[5]) F. de Jager, Les limites théoriques de la transmission en cas de niveau de bruit élevé, pour différents systèmes de modulation continue et de modulation codée, Onde électrique **34**, 675-682, 1954. The ratio C/E_0 used in this paper corresponds to the intrinsic signal-to-noise ratio $(S/N)_i$ employed in the present article. — The remarks made here concerning limiting curves apply only to direct links, for when communication takes place via one or more relay stations the noise arising from repeated re-transmission cumulates in a quite different manner in the four systems. In this respect pulse-code systems behave much better than the others; the shift to the right of the limiting curve when the signal is relayed via one or more intermediate stations is slight in pulse-code modulation, whereas in the other systems it is considerable.

[6]) See for example Philips tech. Rev. **9**, 161-170, 1947/48.

noise ratio is equal to the intrinsic. For $(S/N)_i$ = 17 dB, with single-sideband modulation, S/N is thus also only 17 dB. On the other hand, at very low values of $(S/N)_i$, below about 6 dB, this system seems to give the best results; the others are too much disrupted, and their limiting curves dip below the chain-dotted curve. In practice the improvement given by the four linear systems is only found to become really effective at even higher values of $(S/N)_i$, above about 10 dB.

Here a law that is generally applicable to telecommunications comes into operation. It runs: "The better the quality of a link at relatively low noise powers, the sooner will the link become useless as the noise level increases". This law can also be seen operating in fig. 1, in which results of frequency modulation in different bandwidths can be compared: the curves that rise highest at high $(S/N)_i$ values are the first to fall off when $(S/N)_i$ is reduced. The reader will find a further confirmation of the law at the close of this article, when various non-linear systems are being compared.

Non-linear systems of speech transmission

A signal-to-noise ratio of about 10 dB appears to be necessary in order that a link with single-sideband modulation should be just usable. Now, let us suppose that in certain circumstances (the medium of transmission, the transmitting power and the distance being given) it is impossible to get an $(S/N)_i$ of more than 6 dB. The reception of speech would appear to be scarcely practicable in such conditions. The value quoted means that the speech signal will only have about twice the strength of the noise voltage at times when the speech signal has its maximum value (the maximum transmitting power being used for this); during passages when the strength of the speech signal varies but little from the average it will be entirely swamped by noise.

We have seen that the linear systems can provide no remedy for this state of affairs. It is nevertheless possible at signal-to-noise ratios of 10 to 5 dB, or even lower, to transmit speech sufficiently well for practical purposes by making use of the peculiarities of speech and adapting the apparatus to them. In this way we arrive at systems which, although unsuitable for signals such as music on account of the distortion they involve, are quite capable of conveying speech in a more or less satisfactory manner. We are here concerned with the most important criterion in the transmission of speech, which is not freedom from distortion or fidelity of reproduction, as for music, but *intelligibility*. In the links that we have in mind, distortion of speech is tolerated for the sake of the increased range that is obtained when working with a lower signal-to-noise ratio.

We have already mentioned peak clipping as one such system. In this, the voltage peaks in the speech signal are clipped off by means of a limiter; this allows one to work with a greater average depth of modulation, and thus bring up weak speech signals above the noise.

In other systems which, like peak clipping, were in use prior to 1940, an attempt is made to attenuate noise by interrupting reception with the aid of mechanical or electronic relays during those periods when no information signal is coming through. The make and break times of such relay circuits must naturally correspond closely with the characteristic properties of speech. We shall see from what follows that "Frena" can be regarded as a further step in this direction, being based on our growing familiarity with the characteristics of speech and on the knowledge of what is permissible and not permissible in the treatment of a speech signal.

The Frena system

In the Frena system the frequency component and amplitude component of the speech signal are transmitted separately and recombined on the receiving side (hence the name, which is made up of *fre*quency a*n*d *a*mplitude). The frequency and amplitude channels have to satisfy different sets of requirements with respect to bandwidth and signal-to-noise ratio; in virtue of this, it has proved possible to combine the signals from the two channels in such a way that, for the same total signal power, a much greater insensitivity to interference can be achieved than in the normal case where both the amplitude and the frequency component are transmitted in one channel.

By way of introduction to this system we would refer to the experiments carried out by Licklider [7]) in 1948; these revealed the surprising fact that a speech signal remains reasonably intelligible even when "infinitely" clipped. Infinite clipping involves the removal of all the amplitudes above, say, 0.1% of the peaks in the speech signal. The *amplitude* component is therefore effectively eliminated, the only information transmitted being that regarding the positions of the "zero-crossings" which are associated with the *frequencies* occurring in speech. Needless to say, distortion is very great and the zero-crossings produced by noise in the absence of speech set up in the output an audible disturbance that has the same amplitude as the speech signal; this makes listening to an infinitely clipped signal of this kind

[7]) J. C. R. Licklider and I. Pollack, Effects of differentiation, integration, and infinite peak clipping upon the intelligibility of speech, J. Acoust. Soc. Amer. **20**, 42-51, 1948.

no pleasant matter. However, these experiments clearly illustrate the essential contribution made by the position of the zero-crossings to the information in speech; the *frequencies occurring in the speech waveforms appear to be of far greater importance for intelligibility than their amplitudes*. Confirmation for this thesis has been obtained from experiments in which, *vice versa*, amplitude changes were exactly reproduced while zero-crossing information was eliminated, artificial zero-crossings being derived from a noise generator. When this was done, intelligibility was nil.

The second improvement is that the envelope shape of the single-sideband signal is also transmitted, via a separate channel, so that the clipped signal can be modulated again with the speech envelope at the receiving end. This supplementary signal carries the amplitude information of the speech; it is somewhat analogous to the pilot signals in telephony that are often transmitted along with the speech for various control purposes.

A block diagram of a circuit embodying these principles is shown in *fig. 3*. The audio signal to be

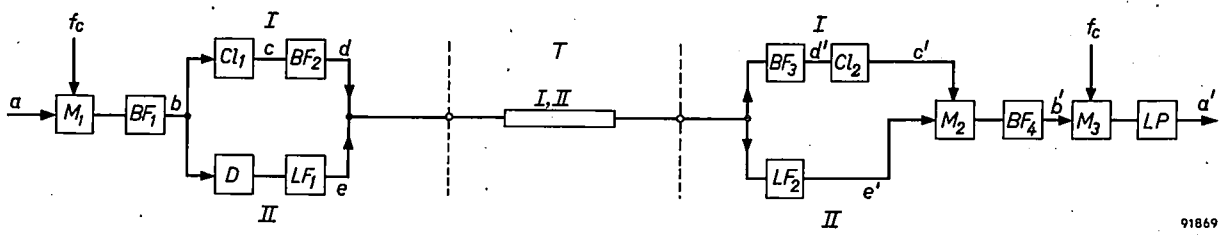


Fig. 3. Block diagram of a link based on the Frena system. The frequency and amplitude components of the speech signal are transmitted via two separate channels, *I* and *II*. M_1, M_2 and M_3 modulators. BF_1 to BF_4 identical band filters having a passband of say 50-53 kc/s. LF_1 and LF_2 identical low-pass filters with cut-off at say 100 c/s. *LP* low-pass filter with cut-off at 3000 c/s. Cl_1 and Cl_2 clippers. *D* detector. *T* transmission path (e.g. carrier-telephony line or radio-relay link). The voltages at points *a*, *b* etc. are shown in fig. 4.

The Frena system is a development based on the Licklider experiments, two essential improvements having been introduced with the idea of keeping distortion within reasonable bounds. In the first place it is not the audio signal itself that undergoes infinite clipping but a single-sideband signal obtained by shifting the audio signal to a higher frequency region at, say, 50 kc/s. In a single-sideband signal of this kind, just as in the audio signal, the information is contained in the position of the zero-crossings as well as in the amplitude. If the amplitude is infinitely clipped, the information still remains available in the zero crossings. The advantage of clipping the single-sideband signal, as against clipping the audio signal, is that the former contains a far greater number of zero-crossings per second and that, in consequence, the displacements of the zero-crossings take place with finer gradation than in the original low-frequency audio signal.

The effect is more or less analogous to the functioning of a normal amplitude detector. When the frequency of a carrier wave is considerably higher than the modulation frequencies imposed upon it, fine detail in the original signal is retained in the shape of the envelope; if, on the other hand, the frequency of the carrier is only slightly higher than the modulation frequency, detection will involve much distortion of the signal. Another way of expressing this is as follows: the clipper, which is essentially a non-linear device, gives rise to a large number of higher harmonics; if the input signal is of low frequency, these harmonics will come within the desired band, but if the input signal is a single-sideband signal, they will fall far outside the band and will therefore be easy to filter out.

transmitted (*a* in *fig. 4*) is impressed on a relatively high carrier-wave frequency f_c, the carrier-wave and one of the sidebands then being suppressed with the aid of band filter BF_1. The single-sideband signal *b* thus arising is infinitely clipped in Cl_1, leaving a square wave signal *c* with zero-crossings of changing phase. The subsequent band filter BF_2, which is similar to BF_1 and whose bandwidth is therefore equal to that of the original audio signal, turns *c* into an approximately sinusoidal signal *d*, which is the one finally transmitted. At the same time the envelope of the single-sideband signal *b* is detected in detector *D*; the signal *e* thus arising can then pass via low-pass filter LF_1 and transmission channel *T* to the receiving end. The transmission path between transmitter and receiver thus comprises two channels, a frequency channel *I* for transmitting a signal (*d*) with no amplitude changes, and an amplitude channel *II* for transmitting a signal (*e*) that does have amplitude changes, but only low-frequency ones.

On the receiving side of channel *I* the first step is that the interference contributed by the transmission path and falling outside the useful frequency band is eliminated by a band filter BF_3, which again is identically similar to BF_1. *d'*, the signal thus obtained, is not of entirely constant amplitude, owing to the interference that remains, but clipper Cl_2 turns it back into a rectangular waveform *c'* which tallies almost exactly with the original clipped signal *c*. If now *c'* is modulated (in M_2) with the envelope

or "pilot" signal e' that has been received from channel *II* and filtered through LF_2, and if the higher harmonics are suppressed by band filter BF_4, a signal b' is obtained that is a very good copy of the original single-sideband signal b. Finally, b' is demodulated with the frequency f_c in the usual manner; thus, from the output of low-pass filter LP, a low-frequency signal a' is obtained that tallies almost exactly with the original signal a. If the frequency bands used in the transmission path (the pass-bands of BF_2 and LF_1) had had unlimited width, an exact copy of the desired signal would have been obtained. The fact that the bandwidths are both limited results in some distortion, but this does not prejudice the intelligibility.

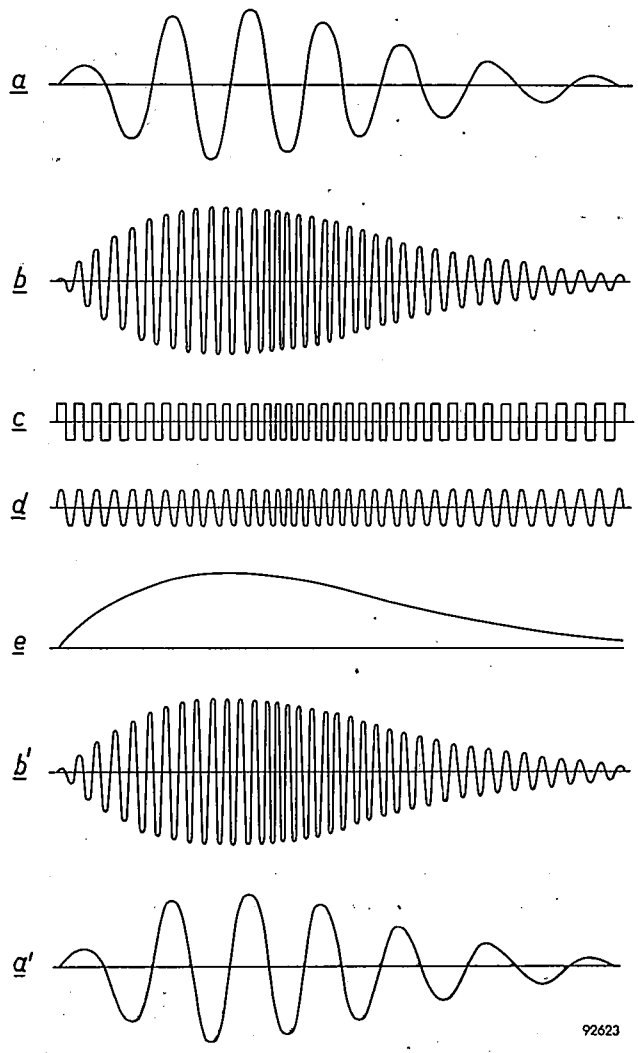


Fig. 4. Shapes of voltages at the points indicated in fig. 3.
a) Speech signal varying both in amplitude and frequency.
b) Single-sideband signal obtained by modulating the carrier wave f_c with the signal *a*.
c) Signal *b* after "infinite" peak clipping; this signal contains only the *frequency* component of the speech.
d) Signal *c* after the higher harmonics have been filtered out.
e) Signal *b* after detection; *e* contains only the *amplitude* component of the speech.
d and *e* are the signals actually transmitted. On being received, *d* is filtered and again "infinitely" clipped, resulting in d'.
b') The signal d' is modulated with e', the received amplitude signal, which has also been filtered, giving b'.
a') The demodulation and filtering of b' produce a nearly identical copy of the original speech signal.

In practice the higher frequencies of the speech signal are given more amplification than the lower frequencies before applying the speech signal to the input of the Frena system (pre-emphasis); at the output of the receiving side the high frequencies are attenuated with respect to the lower ones (de-emphasis), so as to bring about an overall equalization. The aim of the pre-emphasis is to strengthen frequency components in the upper portion of the speech spectrum, which are generally weak, in such a way that they are able to exercise as much effect on the clipping operation as the lower frequencies do. The result is that the overall amount of distortion is lessened. Licklider found the same effect for the original low-frequency speech signals when subjected to infinite clipping.

Noise-suppression properties of the Frena system

The reasons why the Frena system has good noise-suppressing properties can be seen as follows. Firstly, the frequency channel has maximum modulation at all times, even when the speech signal is weak; this allows of a relatively high noise level. Secondly, relatively little noise is received from the amplitude channel, because it is possible to narrow down its bandwidth considerably (see below) [8]). Added to this is the important psychological effect of receiving noise that is modulated in the rhythm of the speech: when the speech signals are faint, little noise is heard at the output of the receiver, and when they are strong the noise is louder in proportion. This makes listening to speech signals received at high noise level much less tiring than in single-sideband transmission, where the noise is of constant strength, making it much more difficult to decide whether one is listening to speech or merely to noise.

The effect of the presence of noise in the frequency channel is illustrated in *fig. 5*. This shows the results of intelligibility tests using logatoms, these being meaningless sound-symbols, each formed according to the pattern consonant(s)-vowel-consonant(s).

[8]) A method of "infinite" clipping of speech in a single-sideband signal corresponding to the frequency channel in our system has also been described by P. Marcou and J. Daguet, Une nouvelle méthode de transmission de la parole, Proc. 3rd Symp. Information Theory, London 1955. However, in the process described therein, the amplitude information is either not transmitted at all, or it is transmitted via a channel not having a narrow band. It is therefore not possible to work with such low signal-to-noise ratios as in our system.

Noise of variable strength was added to the signal in the frequency channel and the percentage intelligibility of the speech sounds was measured as a function of the signal-to-noise ratio in that channel. No noise was injected into the amplitude channel (which in these experiments had a bandwidth of 250 c/s). For the sake of comparison, intelligibility over a normal single-sideband transmission path is shown in the graph by a chain-dotted line. In view of the small number of persons taking part (four) and the somewhat limited number of logatoms used (700 per curve), the results of the tests can only be taken as a rough guide. However, they do demonstrate the characteristic way in which Frena differs from single-sideband modulation.

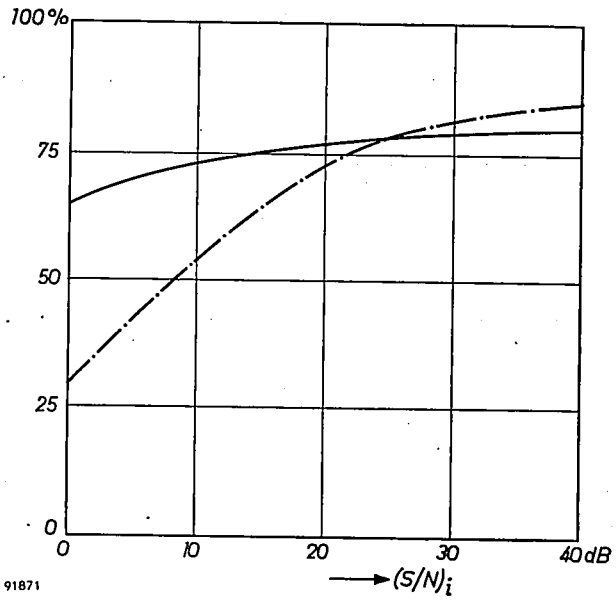


Fig. 5. Intelligibility of speech transmitted by the Frena system (i.e. percentage of sounds, embodied in logatoms, that were correctly recognized), when noise of variable strength was added to the frequency channel. There was still a reasonable degree of intelligibility even when noise and signal had the same power, i.e. when $(S/N)_i = 0$ dB. The chain-dotted curve relates to normal single-sideband modulation.

It will be seen from fig. 5 that when the signal-to-noise ratio is high, the degree of intelligibility given by the Frena system is somewhat inferior to that from a single-sideband system. This is a consequence of distortion which occurs in the Frena system even when noise is absent and which is due to the reduction in bandwidth of the amplitude channel. At lower signal-to-noise ratios, intelligibility falls off much less rapidly than it does with single-sideband modulation. Thus speech continues to be reasonably intelligible even when the signal-to-noise ratio in the frequency channel is 0 dB (that is, when signal and noise have equal power). One point worthy of notice is that persons taking part in the tests remarked on the fatigue of listening to high noise levels in single-sideband modulation, but made no such comments in the case of Frena modulation.

These results suggest the possibility of working, with speech at least, in the region to the left of the curve in fig. 2, which is by definition unattainable for linear systems. To investigate this possibility it is necessary that the amplitude channel, which also requires part of the total signal power, should be taken into account. Experiment has shown that the noise voltage has to be kept considerably lower in this second channel than is necessary in the frequency channel; a signal-to-noise ratio of 20 to 25 dB was found to be desirable. As against this, the bandwidth of this channel can be drastically reduced, even down to a width of only about 50 c/s [9]). This reduction in bandwidth lends a certain coarseness to speech sounds, without however prejudicing intelligibility to any marked extent. It is not desirable to continue the reduction any further, for then the difference in delay times of the frequency and amplitude signals (due to the differing bandwidth of the filters) becomes excessive.

Now, the minimum signal-to-noise ratios at which Frena is workable are also closely connected with the manner in which information is transmitted in the amplitude channel. It can be transmitted by means of amplitude modulation, frequency modulation or code modulation. It will be seen that these modulation methods, in the order given, allow higher and higher noise levels to be tolerated and hence greater and greater ranges to be covered. The manner of transmitting information in the frequency channel (that is, as a single-sideband signal of constant amplitude) always remains the same, however. It is enough for intelligibility if the signal level in that channel be as high as the noise level (see fig. 5), but for the sake of reducing distortion it is desirable that the former should be somewhat higher than the latter, say by 3 dB.

[9]) The fact that when this is done the fundamental tones of the speech for both men and women (100-300 c/s) continue to be properly heard at the receiving end must be attributed to the action of the frequency channel. It is true that this channel does not pass the fundamental itself (nor does a normal telephonic channel), but it does pass a series of its harmonics. These create in the signal a certain periodicity which, according to the "residue" theory, provides us with the information about the fundamental (see J. F. Schouten, The perception of pitch, Philips tech. Rev. 5, 286-294, 1940). In contrast to normal telephony, where the "residue" theory is likewise capable of explaining why we hear the fundamental, it is not possible for the said periodicity to be present in the *amplitude* of our frequency signal, as this has had its amplitude variations completely eliminated; this periodicity can only be present in the phasing of the zero-crossings. Our organ of hearing is evidently capable of discerning information about the fundamental in this phase modulation, and it is thanks to this that the bandwidth of the Frena amplitude channel can be cut down so drastically.

Transmission of information in the amplitude channel

If the amplitude information is transmitted as in fig. 3, channel *II* being, if desired, displaced to a higher frequency range, it is found that there is only a relatively small overall increase in the noise level that can be tolerated. The reason is that the transmitting power necessary for the amplitude channel is relatively high. Suppose for a moment that the frequency channel has a bandwidth of 3000 c/s and a signal-to-noise ratio of 3 dB; then, for an amplitude channel with a bandwidth of 100 c/s and a transmitting power equal to that for the frequency channel, the signal-to-noise ratio will be higher by $10 \log 3000/100 \approx 15$ dB, thus becoming 18 dB. In view of what was said above, this is a little on the low side; it will therefore be better to make the ratio about 3 dB higher, that is, to double the power in the amplitude channel. This makes the total transmitting power three times (i.e. about 5 dB) higher than that in the frequency channel. Transmitting with the same total power and under the same noise conditions by the normal method of single-sideband modulation, we would thus obtain a signal-to-noise ratio of 8 dB. Since a signal-to-noise ratio of 10 dB is required in a normal single-sideband link, the gain given by the Frena system only amounts to 2 dB, or a drop of 37% in the transmitting power required. Although in certain cases this reduction might be of some value, it is desirable that more economic methods be sought for transmitting information through the amplitude channel. For this purpose the employment may be considered of one of the linear noise-suppressing systems we discussed earlier — frequency modulation, for example.

In contrast to the situation in the frequency channel, where the signal-to-noise ratio (which for the present purpose takes the place of the intrinsic signal-to-noise ratio, see fig. 2) is so low that there would be no advantage in using frequency modulation, its employment in the narrow amplitude channel might well be of real utility. As we saw above, an amplitude channel with the same transmitting power as the frequency channel, and with a bandwidth of 100 c/s, has an intrinsic signal-to-noise ratio of 18 dB. This is a value which, according to fig. 2, is well within the region where the use of frequency modulation offers the possibility of considerable improvement. It is in fact found that using frequency modulation with $k = 4$ (see fig. 1), so that the frequency band extends over 400 c/s, it is possible to obtain a signal-to-noise ratio of about 30 dB in the amplitude channel. (Because the envelope in the amplitude channel can only have positive values, the frequency deviation in the FM signal has to occur only in one direction.) The use of frequency modulation thus has the consequence that the amplitude channel requires no more power than the frequency channel. In this way it becomes practicable to transmit information with a total power and under conditions of noise such that in a normal single-sideband link, the signal-to-noise ratio would only be 6 dB.

A further investigation shows that frequency modulation is not yet the best possible method of transmitting information in the amplitude channel. A special system of code modulation will now be described that allows a further gain of 2 dB in permissible noise.

"Frenac", a Frena system using a coded amplitude signal

The reason why a good signal-to-noise ratio is so desirable in the amplitude channel is not so much in order that the envelope shape should be faithfully reproduced, but in order to avoid an undesirable combination effect of the noise in this channel with that in the frequency channel. The needlessness of faithfully transmitting the amplitude signal is clear from the fact that all its peaks can be clipped off with impunity, down to a level of 10 or 15 dB below the topmost; the quality and intelligibility of the speech signal are thereby but little impaired. If, however, the noise in the amplitude channel is increased when the noise level in the frequency channel is already very high, then there is a sharp decrease in intelligibility. The reason is that noise in the frequency channel is passed on to the output even at times when no speech signal is present, between two words or two phonemes, for example: this is a consequence of the fact that the modulator output at the receiving end is proportional to the complete signal in the amplitude channel, including the noise. It is these confusing packets of noise that spoil the intelligibility, because they often have the same character as the start of certain phonemes. To save the listener from continually having to decide whether a signal is to be interpreted as part of a phoneme or as noise, the best course is to suppress these noise components entirely when no speech information is being passed.

Realizing this, and conscious that it was possible, as stated above, to clip the signal carrying the amplitude component very considerably, we decided to replace the amplitude signal by a much simpler pilot signal: the amplitude changes (curve *e* in *fig. 6*) that are detected on the transmitting side are reduced to a square waveform (*f*) that merely indicates whether speech is present or not. The

criterion taken for the presence of speech is whether the amplitude is greater by a certain amount than the microphone noise. The square waveform is transmitted as a series of short wave-trains of constant amplitude (g). Although noise causes the envelope to take on a somewhat erratic shape (f'), the detection of the received signal, which takes place by means of an electronic relay reacting to half the peak voltage, produces (f'') a fairly good approximation to (or reconstruction of) the original square wave signal. This, in fact, is the method of reconstruction by virtue of which pulse-code modulation systems in general are enabled to

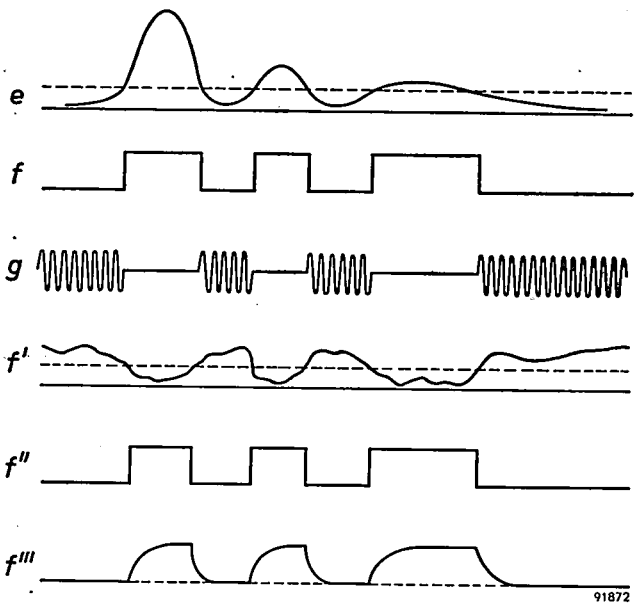


Fig. 6. In the Frenac system the amplitude component (e) of the speech signal is reduced to a square waveform (f) that is transmitted in the form of short wave-trains (g). Detection at the receiving end yields a signal (f') that has been deformed by interference; but this can be used to reconstruct the square waveform (f'') which, after rounding off (f'''), serves to switch the frequency channel on and off at the appropriate instants.

suppress noise. After its sharp edges have been rounded off (f'''), the reconstructed pilot signal can be applied to the first modulator in the receiver.

If the system has a bandwidth of about 100 c/s, the electronic relay can respond rapidly enough for switching to take place even between the successive phonemes of a quickly pronounced syllable.

This, the Frenac system using a coded amplitude signal [10]), provides about the same degree of intelligibility as is obtained from a non-clipped amplitude signal with a high signal-to-noise ratio, though the distortion is somewhat higher. Its great advantage is, however, that absolutely no extra power is required for transmitting the (rudimentary) amplitude component. As may be seen from fig. 6, the pilot signal has a certain constant value when speech is absent, and is zero when it is present. This amounts to saying that the frequency signal and the amplitude signal are transmitted in turn, and the obvious course is to transmit them at the same amplitude. If a signal-to-noise ratio of 4 dB exists in the frequency channel (this having a bandwidth of 3000 c/s), the amplitude channel (of bandwidth 100 c/s) will then be found to have a signal-to-noise ratio 15 dB higher, or about 19 dB. The noise is then sufficiently weak to allow the pilot and interfering signals to be separated in the manner described above (cf. f' and f'' in fig. 6). In many cases the carrier of the single-sideband signal in the frequency channel can profitably be made to act as pilot signal. One advantage of so doing is that there is less trouble from intermodulation between the two channels.

The block diagram of the transmitter and receiver arrived at in this way is given in *fig.* 7, the situation being that at the instant when a phoneme is being transmitted. Relay R responds even to a low level of speech, after detection by D_1. R causes S_1 to switch over from the carrier wave f_c to the clipped single-sideband signal. By using an electronic relay, switching can of course take place very rapidly. The bandwidth required for the transmission path comprises the range of frequencies of the carrier wave and one complete sideband (0-100 c/s for the pilot signal and, say, 300-3300 c/s for the frequency signal).

At the receiving end, the presence or absence of a carrier wave voltage is detected by means of a narrow band filter DF_2 and a detector D_2, allowing the conclusion to be drawn as to whether phonemes are being received or not. If not, the path to the output is broken by the action of relay R' (preferably an electronic relay). Unlike relay R at the transmitter, which responds even to very weak speech signals, the relay on the receiving side is so adjusted that it makes and breaks at approximately half the value of the normal voltage amplitude coming from the detector, in order that it may be able to eliminate noise and interference signals below this halfway value. Noise must therefore rise above a high threshold value before it can cause interference in the amplitude channel; at the noise level assumed above, it is hardly likely that this threshold value will be exceeded even by peak noise voltages, in spite of the fact that the entire link is operating at a signal-to-noise ratio of only 4 dB.

[10]) Reduction of the amplitude signal to a simple choice between yes and no, as here described, is to be regarded as the simplest possible way of coding. It is possible in principle to elaborate the Frenac system by using a multi-level code to carry the amplitude content.

In practice the transmitting end of fig. 7 can be simplified still further by basing it on the scheme shown in *fig. 8*. Together with the single-sideband signal b, a carrier-wave voltage of relatively low amplitude is applied to the input of clipper Cl_1. above-mentioned value of 4 dB. The speech sounded somewhat "grainy" in consequence of the speech-modulated noise, but the degree of intelligibility was highly satisfactory and listening was not tiring. It was possible to alter the input level of speech at the

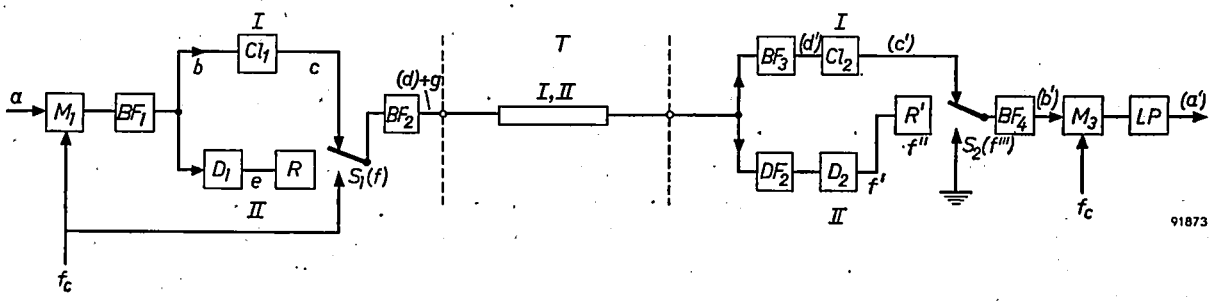


Fig. 7. Block diagram of a link based on the Frenac system. On the transmitting side the detected amplitude changes in the speech signal cause relay R to make or break; this amounts to saying that the said amplitude changes undergo simplification into a square wave signal. S_1 operates in the rhythm of the square wave voltage, alternately passing the frequency signal (c) or the carrier wave f_c on to the transmission channel T. At the receiving end DF_2 filters out the carrier wave and, as a result of detection in D_2, the square wave signal is recovered. The square wave signal, via relay R' and its contacts S_2 (which automatically provide the rounding-off referred to in the text, producing signal-shape f'''), causes the frequency channel to be connected to the demodulator M_3 during the intervals when a speech signal is present.

The remaining letter symbols have the same meaning as in fig. 3. For the voltage wave-forms at the points marked a etc., see figs. 4 and 6. At the points where the letters a-d are enclosed in brackets, the signal in question is present only intermittently.

Providing the amplitude of this additional voltage is low with respect to the normal speech amplitude, there will be but little change in the clipped single-sideband signal while speech is present; on the other hand, when speech is absent, a carrier-wave frequency of the same amplitude will automatically arise at the output of the clipper.

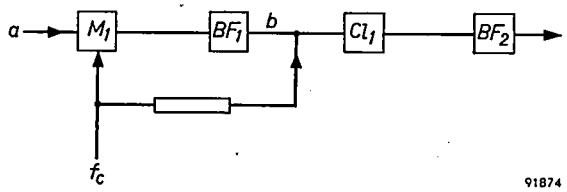


Fig. 8. Block diagram of a simplified circuit at the transmitting end of a Frenac system.

In discussing the fig. 3 circuit we saw that it was of advantage to employ pre-emphasis at the transmitter and de-emphasis at the receiving end. Similarly, in a circuit based on fig. 7 or fig. 8, it is found to be advantageous to carry out such an equalization in front of and behind the Frenac systems and this can be done by means of differentiating and integrating networks respectively. The output signals a', which have been brought down to a constant level as a result of the clipping procedure, thereby acquire a frequency spectrum that corresponds more closely to the average spectrum of speech.

Employment of the Frenac system over a high-frequency transmission path showed that it was possible to work at a signal-to-noise ratio having the transmitter by about 15 dB in all without intelligibility suffering too much from the increasing distortion as the signal strength rose. When the noise level was raised to the point where the signal-to-noise ratio became 0 dB (so that the noise and the signal in the transmission path had the same r.m.s. value), the link proved to be still serviceable providing one spoke slowly. There was indeed some loss of intelligibility, in consequence of which it was no longer possible to alter the input level so widely: it was permissible for this level to vary only over about 5 dB in all. Should the noise level be raised even more, so that it rises above the signal level, the degree of intelligibility falls off sharply; this is because the amplitude channel as well as the frequency channel then suffers from serious interference by the noise.

Comparison of the various systems with respect to the minimum required signal-to-noise ratio, gives us the following figures.

	Minimum required signal/noise power ratio, $(S/N)_i$
Single-sideband modulation	10 (10 dB)
Frena with AM in amplitude channel	6 (8 dB)
Frena with FM in amplitude channel	4 (6 dB)
Frenac (using coded amplitude signal)	2.5 (4 dB)

From this table it is seen that Frena systems make it possible to transmit speech at signal-to-noise ratios so low that the noise-suppressing action of familiar linear systems such as frequency modulation would, in the same conditions, have long ceased before to be effective; at the same time there is a big gain in comparison with normal single-sideband modulation. All this is a consequence of the fact that the systems in question have to a great extent been adapted to the characteristic properties of speech and the human ear. That these systems cannot be considered for transmitting other kinds of signal, such as music, detracts little from their usefulness; after all, many telecommunication links — most of them, in fact — are designed exclusively for conveying speech. The Frenac system, which allows speech to be passed through very strong noise, has the further property that the complete signal as radiated has constant amplitude. This offers an opportunity of using transmitters with tubes under class C operation, if so desired, and thus of obtaining high efficiency. Finally the presence of the pilot signal when phonemes are absent (and that means most of the time) is of interest in cases where it is desired to use automatic frequency control on the receiver.

Although Frenac is being put forward here as an attempt to get the most effective conveyance of speech with a certain, just adequate quality in the presence of a high level of interference, the question arises as to whether Frenac is of any use for *improving* quality at intrinsic signal-to-noise ratios that are not extremely poor, 10 dB, say. This train of thought brings us back to our account of the linear systems, where we dealt with the matter of improving the effective signal-to-noise ratio for a given intrinsic signal-to-noise ratio. In discussing the Frena and Frenac systems we have made no use of the term "effective signal-to-noise ratio", since this quantity (even should its definition be modified) cannot serve as a yardstick for quality in speech links of this kind, on account of the distortion they involve. Indeed, in the first instance we had no need of such a yardstick, for we assumed a substantially fixed minimum quality, and merely inquired how far the intrinsic signal-to-noise ratio required to give that quality might be reduced. There is, however, a way of giving a numerical rating to the quality of speech links that involve distortion: this is by listening tests that are carried out according to a procedure suggested by the C.C.I.F. (Comité Consultatif International de Téléphonie) [11]). The rating thus obtained takes account of various criteria such as intelligibility, distortion, ease of listening, and so on. By means of such tests we have been able to establish quantitatively that the quality of Frenac can indeed be considerably improved at the above-mentioned not unduly bad signal-to-noise ratio of 10 dB. The improvement is effected by not allowing the pilot signal to switch the frequency channel on and off with complete discontinuity. This constitutes as it were a step backward towards the original idea of Frena, but without there being any question of linearity in the transmission of the amplitude component. This modified Frenac system exhibited a clear improvement in quality not only as compared with the pure Frenac system but also as compared with normal single-sideband modulation, either with or without clipping. Investigation along these lines is still proceeding.

[11]) Draft report of 4th study group, C.C.I.F., Geneva, November 1956.

Summary. For the transmission of speech by means of single-sideband modulation, a signal-to-noise ratio of at least 10 dB is necessary. Systems such as FM and PCM make it possible to improve the signal-to-noise ratio in a communications link by giving the transmission path a wider frequency band. However, these "linear" interference-suppressing systems are only effective where the "intrinsic" signal-to-noise ratio of the link is not too low. The intrinsic ratio must indeed be higher than about 10 dB and hence, at the low signal-to-noise ratios mentioned above, the said linear systems offer no advantage over single-sideband modulation. In contrast, non-linear systems that have been specially designed for the transmission of speech make it possible to pass intelligible speech signals at even lower signal-to-noise ratios. Noise up to a certain level can be suppressed by permitting a degree of distortion to take place that does not unduly prejudice the intelligibility. This principle has been put into practice in the "Frena" system developed in Eindhoven, in the following manner: the frequency component and amplitude component of speech are conveyed via two separate channels, the first being transmitted as a single-sideband signal in a frequency band about 3000 c/s in width, the second in a band about 100 c/s in width. In the frequency channel a signal-to-noise ratio as low as 3 dB is quite adequate. However, owing to the fairly high power required for transmitting the amplitude component, the necessary minimum signal-to-noise ratio for the whole system remains at about 8 dB when the amplitude component is transmitted as an AM signal. If FM is employed, the necessary minimum is brought down to about 6 dB. It can be lowered still more by reducing the amplitude information to a "pilot" signal which merely indicates at each instant whether a speech sound is present or not (the "Frenac" system). The apparatus can then be designed in such a way that the pilot signal requires no additional power at all; as the noise is suppressed at the most critical times, it interferes so little with the speech signal that the latter can be received intelligibly even at a signal-to-noise ratio of only 4 dB.

Although the basis of the above developments was minimum quality, the endeavour being to preserve this quality at ever-worsening signal-to-noise ratios, efforts are at present being directed to develop further variants of Frena and Frenac giving reception of better quality at not unduly bad signal-to-noise ratios, at 10 dB for example.

CHLADNI'S FIGURES ON THE VIBRATING CAPACITOR OF A SYNCHROCYCLOTRON

621.384.611.1:621.376.32

In a synchrocyclotron the frequency of the R.F. field accelerating the particles is periodically varied. In the larger cyclotrons this frequency modulation is effected by varying the capacitance in the oscillatory circuit, usually by means of a rotating capacitor [1]). With the new 600 MeV synchrocyclotron of the CERN (Conseil Européen pour la Recherche Nucléaire) at Geneva, which is to be completed in the course of this year, a vibrating capacitor will be used to bring about this capacitance variation. The equipment for this purpose has been developed by CERN and the Philips Research Laboratories of Eindhoven, working in cooperation.

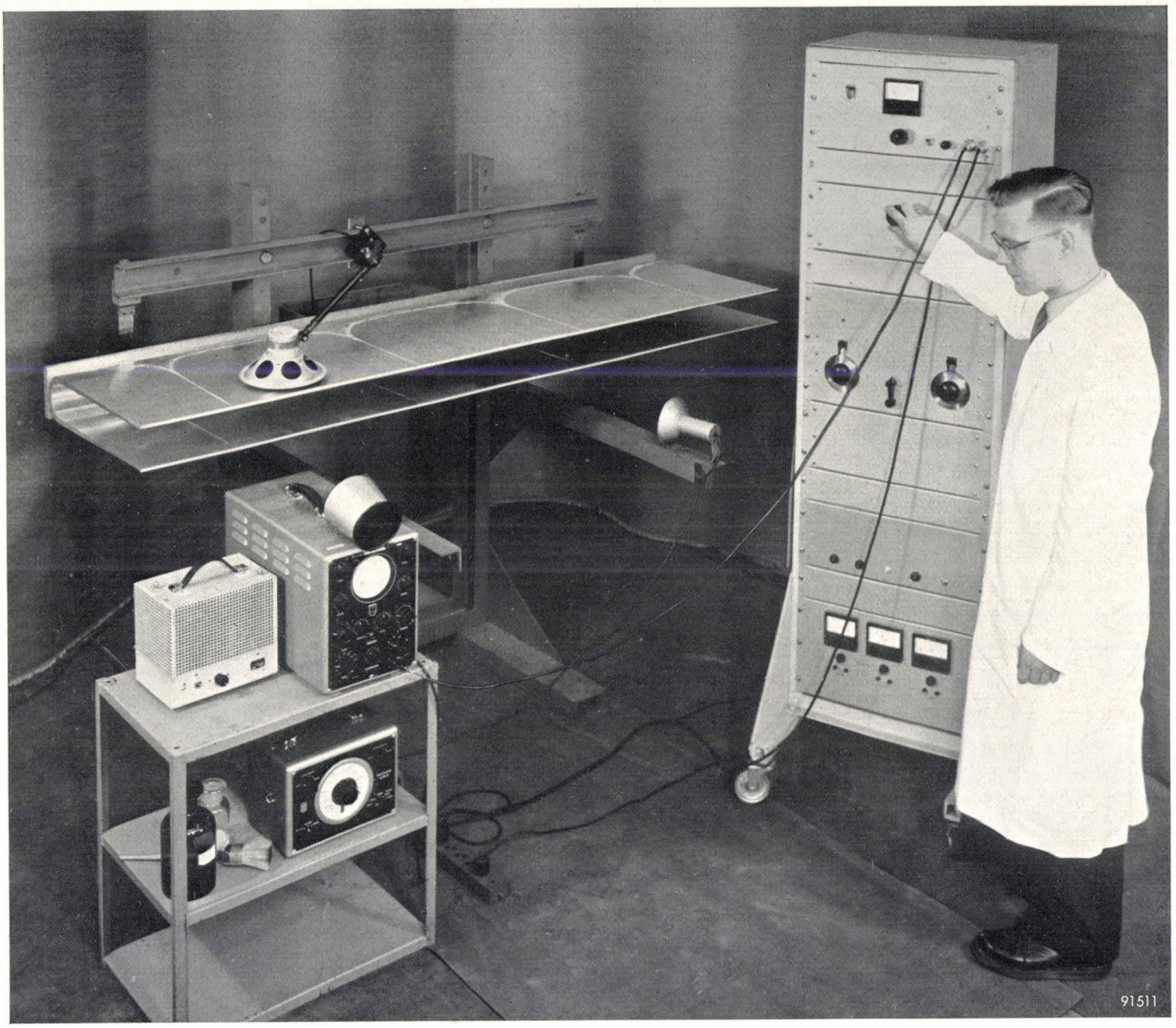


Fig. 1. Arrangement for investigating the resonances of the vibrator. The loudspeaker, mounted close to but not touching the upper blade of the vibrator, is driven by the audio generator at the right. The sinusoidal form of the supply voltage is checked with the oscilloscope (left), whilst the movement of the blades can be further analyzed with a stroboscope, the lamp of which can be seen on the supporting beam to the right and in front of the vibrator. When the vibrator is in certain states of resonance, powder sprinkled on the blades collects upon the nodal lines, thus making the vibration pattern visible, in a manner similar to the "sound-figures" made by Chladni of vibrating plates and the soundboards of musical instruments.

The pattern visible here corresponds to 110 c/s.

The vibrating capacitor consists of a 2 metres wide aluminium vibrator having a cross-section like

[1]) See for example F. A. Heyn, The synchrocyclotron at Amsterdam, II. The oscillator and the modulator, Philips tech. Rev. **12**, 247-256, 1950/51.

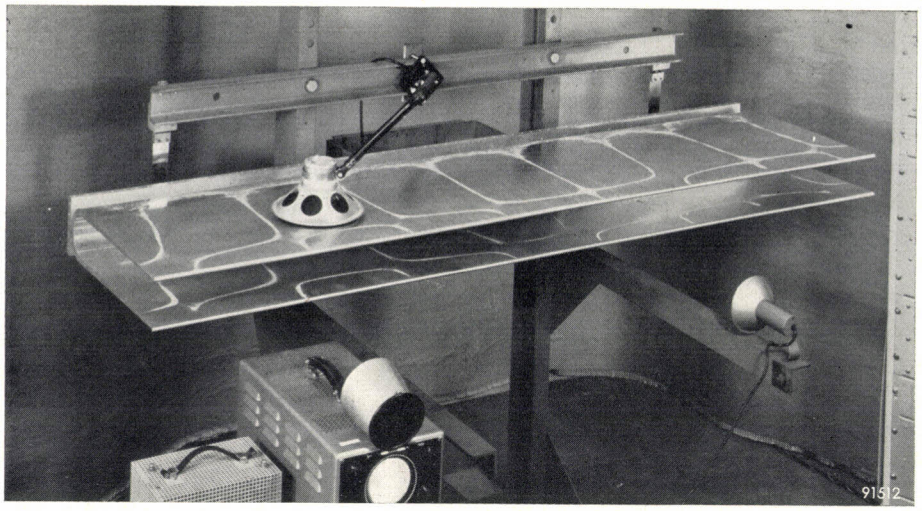


Fig. 2. Vibration pattern at 540 c/s.

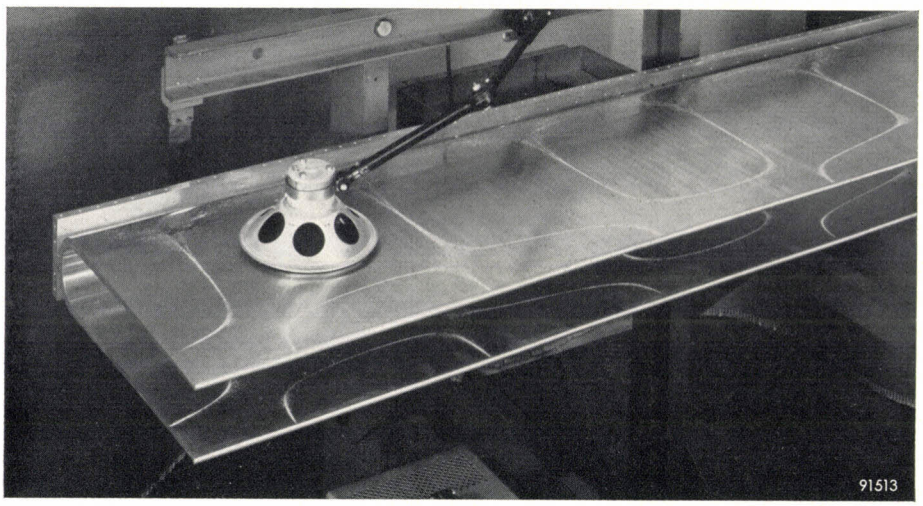


Fig. 3. Vibration pattern at 380 c/s.

a tuning-fork. A fixed electrode is mounted between the blades. The vibrator has a fundamental frequency of 55 c/s, the required amplitude being 12.5 mm. It is excited by setting up a longitudinal vibration of the above frequency in the "stem" of the fork. We shall not enter here into details regarding the excitation and the arrangements for controlling the amplitude.

For its proper operation it is necessary for the tuning-fork to vibrate in its fundamental mode only, this being when the two blades close up and move apart in their entirety and no nodal lines occur. To make sure that no interfering resonances of the vibrator can occur while the synchrocyclotron is in operation, an investigation was made into the vibration patterns and frequencies of the overtones of the vibrator. The photographs in *figs. 1-3* show the experimental arrangement and some striking Chladni's figures relating to some of the many possible (and undesired) vibration modes of the vibrator. For these experiments the vibrator was acoustically excited at its various resonance frequencies with the aid of a small loudspeaker fed from an audio signal generator (range 27-1800 c/s) with precision tuning. A fine powder was then sprinkled onto the blades to produce the Chladni's figures. A stroboscope made it possible to analyze the phase difference in the motion of the various parts of the blades.

This investigation revealed that at the rated frequency there was an undesirable resonance, having the vibration pattern shown in fig. 1: its resonance frequency happened to be almost exactly twice the fundamental frequency. By fitting on both blades of the tuning-fork a stiffening rib running along the entire width at some distance from the front edge, the resonance was shifted to a higher frequency, where it was no longer troublesome. Owing to the stiffening of the blades, the resonance frequencies of all other modes of vibration were, of course, also shifted, but no new unwelcome coincidences between such frequencies and multiples of the fundamental then occurred.

B. BOLLÉE.

FUNDAMENTALS OF COLOUR TELEVISION

by F. W. de VRIJER.

621.397.2

In recent years the colour printing of illustrations in magazines and books, and the colour film in the cinema and in amateur photography, have attained a high degree of perfection. They have, as it were, opened the general public's eyes to the world of colour. Colour television has also made considerable strides, as evidenced, for example, by the regular colour television transmissions in the United States. In view of the high costs at present involved, colour transmissions are being made on a relatively small scale; it is not to be expected, therefore, that colour television will, for the time being, be introduced in Europe other than experimentally.

The article below is an introduction to a series of articles to be printed in this Review describing some of the work being done by Philips in the field of colour television.

One of the fundamentals of colour television is, of course, the science of colour. Various colorimetric concepts, such as the chromaticity diagram (colour triangle), have already been discussed in this Review [1]). We shall begin by briefly recapitulating these concepts.

Colorimetry

The spectral energy distribution of a light source determines the colour of the light. On the other hand, the colour of the light does not uniquely establish the spectral distribution, for different spectral distributions can produce the same colour sensation. It was known to Newton that almost all colours can be imitated by additively mixing light of only three primary colours, i.e. by so combining them that the various types of light impinge simultaneously on the same part of the observer's retina. Additive mixing occurs, for example, when two or more projectors, each with a light beam of a given colour, illuminate the same part of a projection screen. If red, green and blue are taken as the primary colours, almost all colours can be reproduced by additive mixing (the minor reservation contained in the word "almost" will be explained later). By adding green to red, for instance, we get orange; a little more green produces yellow and still more green results in yellowish green. Mixed in the right proportions, the three primary colours can produce white (*fig. 1*).

The converse to additive mixing is subtractive mixing. This arises when specific parts of the spectrum of a light source are attenuated more than others, for example by passing the light through a filter or by making it reflect selectively from a surface. The latter is involved in the mixing of pigments [2]) and is, for example, the basis of modern forms of colour photography. The laws of subtractive mixing are more complicated than those of additive mixing (we shall return to this subject later). *Fig. 2* gives an example of the difference in result as against fig. 1, and shows the effects of mixing two and three paints whose colours are complementary to those of the light on which fig. 1 is based.

The primary colours frequently used in colorimetry are the spectral colours given below and defined by their associated wavelengths λ (in vacuo):

$$\left.\begin{array}{ll} \text{Red,} & \lambda = 700.0\ \text{m}\mu \\ \text{Green,} & \lambda = 546.1\ \text{m}\mu \\ \text{Blue,} & \lambda = 435.8\ \text{m}\mu \end{array}\right\} \quad \ldots \quad (1)$$

If the eye is unable to distinguish a given type of light from the additive mixture of B_1 lumens at $\lambda = 700.0$ mμ, B_2 lumens at $\lambda = 546.1$ mμ and B_3 lumens at $\lambda = 435.8$ mμ, we then say that the light concerned has the colour coordinates B_1, B_2 and B_3 in the system based on the above-mentioned primary colours. *Fig. 3* gives the result of measurements — averaged over a large number of observers — of the primary-colour coordinates of 1 lumen of any spectral colour. It can be seen that negative values also occur. For example, if B_1 is negative while B_2 and B_3 are positive, it means that the additive mixture of the original light and $(-B_1)$ lumens of $\lambda = 700.0$ mμ, appears the same to the eye as the additive mixture of B_2 lumens of $\lambda = 546.1$ mμ and B_3 lumens of $\lambda = 435.8$ mμ. Such types of light, with one or two negative coordinates, cannot therefore be imitated by additively mixing the three selected primary colours.

[1]) See W. de Groot and A. A. Kruithof, Philips tech. Rev. **12**, 137-144, 1950/51. For a fuller treatment see P. J. Bouma, Physical aspects of colour (Philips Technical Library 1947).

[2]) See J. L. H. Jonker and S. Gradstein, Fluorescent pigments as an artistic medium, Philips tech. Rev. **11**, 16-22, 1949/50.

If, instead of one spectral colour, we have light with a given spectral energy distribution $E(\lambda)$, we may write for the coordinates:

$$\left.\begin{aligned} B_1 &= \int \bar{B}_1(\lambda) E(\lambda)\, \mathrm{d}\lambda, \\ B_2 &= \int \bar{B}_2(\lambda) E(\lambda)\, \mathrm{d}\lambda, \\ B_3 &= \int \bar{B}_3(\lambda) E(\lambda)\, \mathrm{d}\lambda. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (2)$$

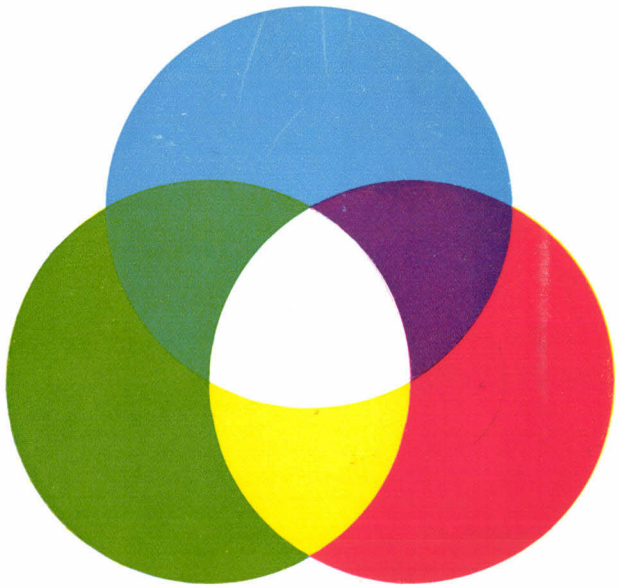


Fig. 1.

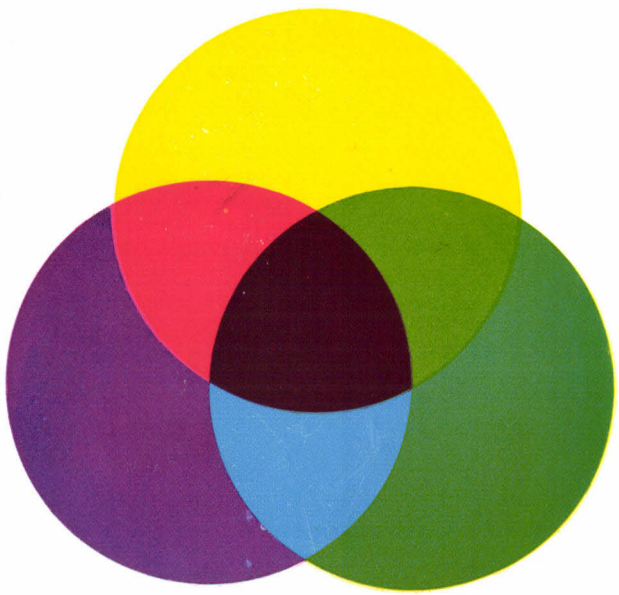


Fig. 2.

Fig. 1. Additive mixing of red, green and blue light. Mixed by twos, these types of light produce yellow, blue-green or purple (magenta). The three types of light mixed in the right proportions produce white.
Fig. 2. Subtractive mixing. A yellow dye mixed with a blue-green dye produces a green mixture, a blue-green and a purple dye produce blue, and a purple and a yellow dye produce red. The three dyes together can produce black.

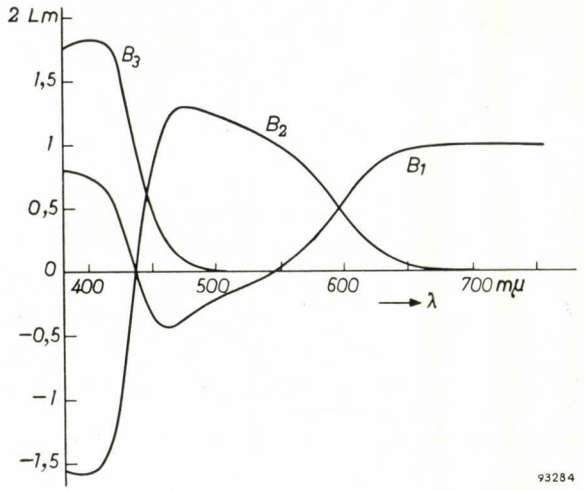


Fig. 3. The colour coordinates B_1, B_2 and B_3 of 1 lumen of the spectral colours with wavelength λ, in the system of primary colours given by (1). ($1\ \mathrm{m}\mu = 10^{-9}$ m.)

The quantities $\bar{B}_1(\lambda)$, $\bar{B}_2(\lambda)$ and $\bar{B}_3(\lambda)$ are represented graphically in *fig. 4*.

It appears, then, that the colour coordinates of a mixture of spectral colours are additive. This additivity applies not only to spectral colours but, as will be clear, it also applies generally to arbitrary spectral distributions.

Two types of light with the same ratio $B_1 : B_2 : B_3$ have the same "type" of colour (or *chromaticity*), but in general they differ in luminance. Luminance has been defined by the Commission Internationale de l'Eclairage (C.I.E.) as the luminous flux

$$K \int V(\lambda)\, E(\lambda)\, \mathrm{d}\lambda \quad . \quad . \quad . \quad . \quad . \quad (3)$$

per unit area and per unit solid angle. In the expression (3) the constant K is approximately 683 lumens/watt, and $V(\lambda)$ is the internationally accepted relative spectral sensitivity of the eye, determined from experiments using large numbers of observers. "Relative" implies here that the maximum of $V(\lambda)$ is made equal to 1; for wavelengths

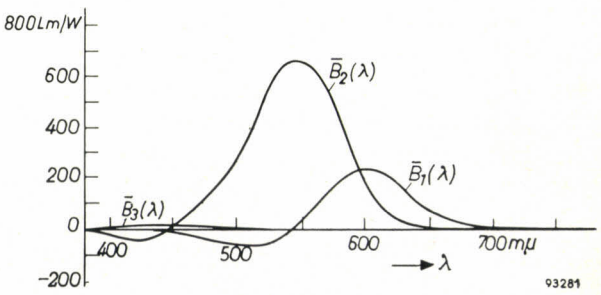


Fig. 4. The quantities $\bar{B}_1(\lambda)$, $\bar{B}_2(\lambda)$ and $\bar{B}_3(\lambda)$ from (2) as a function of λ. This figure differs from fig. 3 in that it is based not on the luminous flux but on the energy flux E. Fig. 4 can be derived from fig. 3 by multiplication by the visual-sensitivity curve.

smaller than 400 mμ and larger than 780 mμ [3]), $V(\lambda)$ is practically zero. The unit of luminance is the candela per m².

From the additivity rule it follows directly that the luminance L of any type of light is

$$L = B_1 + B_2 + B_3 = \int \{\bar{B}_1(\lambda) + \bar{B}_2(\lambda) + \bar{B}_3(\lambda)\} E(\lambda) \mathrm{d}\lambda.$$

Thus, according to (3),

$$KV(\lambda) = \bar{B}_1(\lambda) + \bar{B}_2(\lambda) + \bar{B}_3(\lambda). \quad . \quad . \quad (4)$$

Instead of taking the coordinates B_1, B_2 and B_3 to characterize a type of light, we may also take three mutually independent linear combinations of B_1, B_2 and B_3. The C.I.E. has laid down as the international system of colour coordinates:

$$\left.\begin{aligned} X &= 2.7689\, B_1 + 0.38159\, B_2 + 18.801\, B_3, \\ Y &= \quad B_1 + \quad B_2 + \quad B_3, \\ Z &= \quad 0.01237\, B_2 + 93.060\, B_3. \end{aligned}\right\} \quad . \quad (5)$$

From this it follows that:

$$\left.\begin{aligned} X &= K \int \bar{X}(\lambda)\, E(\lambda)\, \mathrm{d}\lambda, \\ Y &= K \int \bar{Y}(\lambda)\, E(\lambda)\, \mathrm{d}\lambda, \\ Z &= K \int \bar{Z}(\lambda)\, E(\lambda)\, \mathrm{d}\lambda, \end{aligned}\right\} \quad . \quad . \quad . \quad (6)$$

where

$$\begin{aligned} K\bar{X}(\lambda) &= 2.7689\bar{B}_1(\lambda) + 0.38159\bar{B}_2(\lambda) + 18.801\bar{B}_3(\lambda), \\ K\bar{Y}(\lambda) &= \quad \bar{B}_1(\lambda) + \quad \bar{B}_2(\lambda) + \quad \bar{B}_3(\lambda), \\ K\bar{Z}(\lambda) &= \quad 0.01237\bar{B}_2(\lambda) + 93.060\bar{B}_3(\lambda). \end{aligned}$$

(See in this connection page 137 of the article quoted in footnote [1]).) A graphic representation of $\bar{X}(\lambda)$, $\bar{Y}(\lambda)$ and $\bar{Z}(\lambda)$ is given in *fig. 5*. Two advantages of this X, Y, Z system [4]) are:

1) one of the coordinates, viz. Y, is identical with the luminance;
2) negative values of the coordinates do not occur.

With reference to the latter advantage, it may be remarked that not every set of X, Y, Z values, although none are negative, corresponds to an existing type of light. For instance, colours corresponding to points on the coordinate axes (two of the three coordinates being zero) do not exist. This is sometimes expressed by saying that "the primary colours of the X, Y, Z system are not real".

It has already been pointed out that types of light with the same ratio of coordinates have the same chromaticity (type of colour) and merely differ in luminance. Clearly this also applies to the X, Y, Z system. To designate the colour, therefore, two quantities are sufficient. Those frequently used are:

$$x = \frac{X}{X+Y+Z} \text{ and } y = \frac{Y}{X+Y+Z}. \quad . \quad (7)$$

Every existing colour may now be represented by a point in a flat plane, with the coordinates x and y. All these points together fill what is known as the chromaticity diagram or colour triangle (*fig. 6*) [5]).

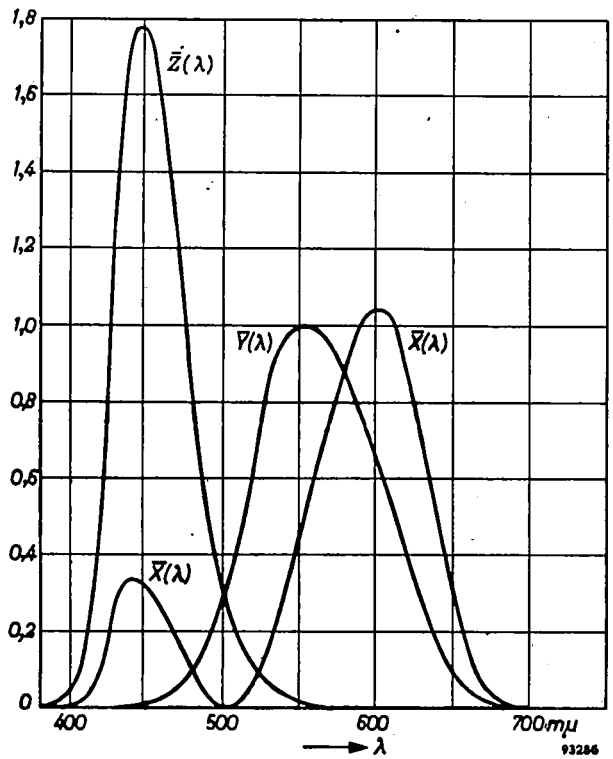


Fig. 5. The quantities $\bar{X}(\lambda)$, $\bar{Y}(\lambda)$ and $\bar{Z}(\lambda)$ from (6) as a function of λ.

The periphery is occupied by the spectral colours and the so-called purple or magenta colours, the middle by white. From the periphery towards white the "saturation" decreases from maximum to zero. The caption of fig. 6 explains the rule for additive mixing (centre-of-gravity rule).

Principles of colour television

Starting from the colorimetric principles described above, we can imagine a colour television system as being realized in the manner shown schematically in *fig. 7*. The light L coming from the scene is split into three components, red, green and blue. These three components each act upon a camera tube (E_r, E_g, E_b), the video signals of which are transmitted via separate cables or radio channels to the associated picture tubes (F_r, F_g, F_b) at the reception

[3]) In addition to the objective concept "luminance" there exists the subjective concept of "brightness". Although brightness is almost synonymous with luminance, it is better not to use the word brightness in a quantitative sense.

[4]) For other advantages (not of importance in colour television) see Chapters 5 and 6 of the book by Bouma mentioned in [1]).

[5]) A reproduction in colour is given in Philips tech. Rev. 12, 141, 1950/51.

end. The latter tubes are provided respectively with a red, a green and a blue luminescent phosphor. Combined in the right way, the three monochromatic pictures produce a single multicoloured picture. (Instead of the projection system as shown here at the receiving end, which is more suitable for the production of large pictures, use is preferably made for home-reception of a special direct-view tube; we shall not discuss this system here.) For the sake of clarity, the amplifiers required for boosting the signal to the picture tubes, and the deflection and synchronization devices, are not shown in fig. 7.

The separation of the incoming light into three components at the transmitting end, as well as the combination of the three pictures at the receiving end, can be effected by means of dichroic mirrors, i.e. mirrors which efficiently reflect a part of the spectrum and transmit the remaining part. An article has recently appeared in this Review, describing the action of these mirrors, which is based on interference, and discussing their manufacture by the deposition of thin layers [6]). This article also describes the principle of combining the three pictures: the red reflecting mirror M_{r2} (fig. 7) transmits green and blue, the blue reflecting mirror M_{b2} transmits green and red, and with their aid the three differently coloured pictures on the receiving tubes (in this case projection tubes) can be projected on to a screen S in accurate register and with little loss of light. This projection system will be dealt with in more detail in a subsequent article.

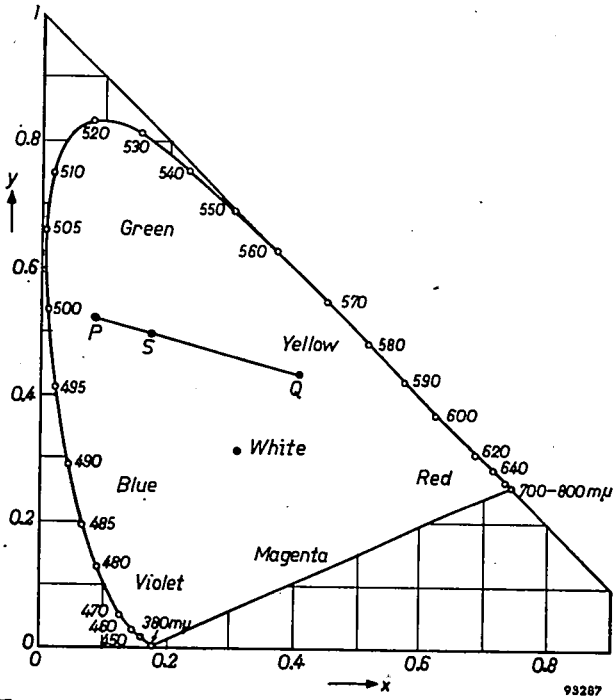


Fig. 6. The colour triangle [5]). The wavelengths of the spectral colours are given around the circumference.

Rule for additive mixing. Assume that Y_P lumens of colour P (with coordinates x_P, y_P in the colour triangle) are mixed with Y_Q lumens of the colour Q (x_Q, y_Q). The international colour coordinates X, Y and Z are then:

	X	Y	Z
Colour P	$X_P = x_P Y_P/y_P$	Y_P	$Z_P = (1-x_P-y_P)\,Y_P/y_P$
Colour Q	$X_Q = x_Q Y_Q/y_Q$	Y_Q	$Z_Q = (1-x_Q-y_Q)\,Y_Q/y_Q$

The coordinates of the colour mixture are found by addition. The result of the mixture has a luminance $Y_P + Y_Q$ and its colour point (S) follows from the centre-of-gravity rule. Imagine that the mass $X_P + Y_P + Z_P$ is fixed at P and the mass $X_Q + Y_Q + Z_Q$ at Q. The centre of gravity S of these masses is then the required colour point, which always lies on the connecting line PQ. Since all types of light consist of a discrete line spectrum or a continuous spectrum, the colour points of all existing types of light lie in the area bounded by the line of the spectral colours and the so-called purple line.

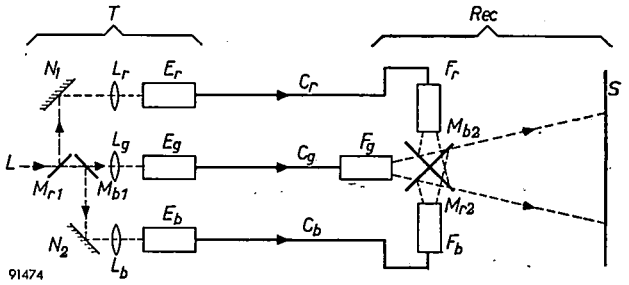


Fig. 7. Schematic representation of a system for colour television.

At the transmission end T, of the light L coming from the scene, the red reflecting mirror M_{r1} separates the red component, which is projected on to the "red" camera tube E_r via the ordinary mirror N_1 and the lens L_r. The blue component reaches the "blue" camera tube E_b via the blue reflecting mirror M_{b1}, the ordinary mirror N_2 and the lens L_b. The green component passes through M_{r1} and M_{b1} and arrives at the "green" camera tube E_g via the lens L_g. Separate cables or radio links C_r, C_g and C_b convey the "red", "green" and "blue" signals to the receiving end *Rec*, where they produce a picture of corresponding colour on picture tubes F_r, F_g and F_b with red, green and blue luminescent phosphors respectively. These picture tubes are each associated with a projector. By means of the dichroic mirrors M_{r2} and M_{b2}, the three pictures are projected in exact register on to the screen S. (Since the beams have a large cross-section in this case, intersecting dichroic mirrors are often used; see [6]).)

Fig. 7 also shows the way in which the incoming light is split into three components, by means of the dichroic mirrors M_{r1} and M_{b1} and the ordinary mirrors N_1 and N_2. This needs no further explanation. It may be added that colour filters can be placed in the three light paths in order to produce a closer approximation to the required spectral sensitivities; we shall return to this later.

On the basis of this extremely simplified colour television system we shall now consider various questions more in detail.

Choice of the primary colours

A colour produced by the additive mixing of two other colours always lies in the colour triangle upon the line connecting these two colours, at a position

[6]) P. M. van Alphen, Philips tech. Rev. **19**, 55-63, 1957/58 (No. 2).

which follows from the centre-of-gravity rule (fig. 6). Starting from three primary colours we can therefore reproduce only the colours that lie within (or on the circumference of) the triangle formed by the colour points of the three selected primary colours. This triangle must accordingly be large enough to contain that part of the colour triangle which is of practical importance. It follows directly from the position of the colours in the colour triangle that, broadly speaking, red, green and blue are the appropriate primary colours.

In order to determine more exactly the primary colours required for a television system it is necessary to take account of some important practical considerations. At the receiving end it is desirable to be able to produce the three primary colours by means of known phosphors, and to do this with an efficiency capable of giving the required luminance without having to apply too great a power to the picture tubes. In the United States these considerations have led to the following choice of primary colours:

	x	y	
Red . . .	0.67	0.33	} . . . (8)
Green . . .	0.21	0.71	
Blue . . .	0.14	0.08	

In *fig. 8* these colour points (which differ from the spectral primary colours given by (1)) are designated by R, G and B. With these primary colours, then, it is possible to produce all colours whose colour point lies within (or on) the triangle RGB. It has been found that almost all the colours occurring in nature, and also those of common dyes (bounded by contour *1* of fig. 8) fall within this triangle. Compared with other well-known colour reproduction processes (colour photography, for example) the part of the colour triangle covered in this case may be regarded as favourable (compare contour *2* in fig. 8 with the triangle RGB). In the following pages we shall base our discussion on the primary colours given by (8).

Reproduction of white

The reproduction of white calls for special attention. The colour point of white, as radiated by a picture tube for monochromatic television, generally lies in the region of so-called standard white C (point C in fig. 8), with the coordinates $x = 0.310$, $y = 0.316$. This colour is generally accepted as a good choice, and it is reasonable, therefore, to take it as the white for colour television too, i.e. a colourless object in the scene should be reproduced by standard white C. A simple calculation shows that in order to produce 1 candela/m² of this white light from the three primary colours (8), we need 0.30 cd/m² red, 0.59 cd/m² green and 0.11 cd/m² blue.

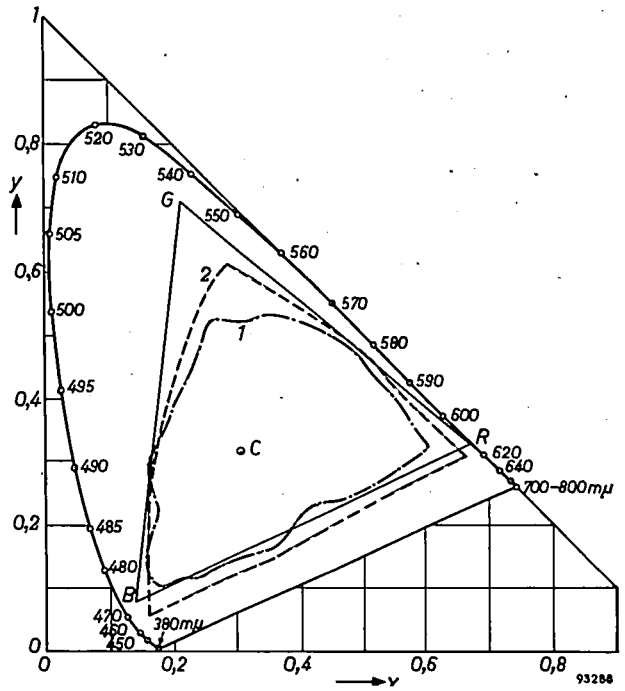


Fig. 8. Colour triangle, with the colour point C of standard white C ($x = 0.310$, $y = 0.316$). R, G and B are the primary colours selected for colour television, with coordinates according to (8). By additive mixing, all colours can be produced whose colour point lies within or on the triangle RGB. This is more favourable than the possibilities in subtractive colour-film processes, in which the contour *2* limits the realizable colours. Within contour *1* lie all the reflection colours occurring in nature and the reflection colours of all known dyes and printing inks.

This calculation is made as follows. For 1 cd/m² white light with $x = 0.310$, $y = 0.316$, $z = 1 - (x + y) = 0.374$, the associated values of X, Y and Z are:

$$X_w = \frac{x}{y} = \frac{0.310}{0.316} = 0.980,$$
$$Y_w = 1.000,$$
$$Z_w = \frac{z}{x} X_w = \frac{0.374}{0.310} \times 0.980 = 1.184.$$

l_r cd/m² red with $x = 0.67$, $y = 0.33$, $z = 0$ gives:

$$\left.\begin{aligned} X_r &= \frac{0.67}{0.33} l_r = 2.03\, l_r, \\ Y_r &= 0.33\, l_r, \\ Z_r &= 0. \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (9a)$$

l_g cd/m² green with $x = 0.21$, $y = 0.71$, $z = 0.08$ gives

$$\left.\begin{aligned} X_g &= \frac{0.21}{0.71} l_g = 0.296\, l_g, \\ Y_g &= l_g, \\ Z_g &= \frac{0.08}{0.71} l_g = 0.113\, l_g. \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (9b)$$

l_b cd/m² blue with $x = 0.14$, $y = 0.08$, $z = 0.78$ gives

$$\left.\begin{aligned} X_b &= \frac{0.14}{0.08} l_b = 1.75\, l_b, \\ Y_b &= l_b, \\ Z_b &= \frac{0.78}{0.08} l_b = 9.75\, l_b. \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (9c)$$

In order to obtain exactly 1 cd/m² standard white C by additively mixing l_r cd/m² of this red, l_g cd/m² of this green and l_b cd/m² of this blue, we must ensure that the following relations exist:

$$X_w = 2.03\ l_r + 0.296\ l_g + 1.75\ l_b = 0.980,$$
$$Y_w = l_r + l_g + l_b = 1.000,$$
$$Z_w = 0.113\ l_g + 9.75\ l_b = 1.184.$$

From these three equations we can solve l_r, l_g and l_b. We find: $l_r = 0.298$ cd/m², $l_g = 0.588$ cd/m², $l_b = 0.114$ cd/m² (rounded off: 0.30, 0.59 and 0.11 cd/m²).

Luminance signal

In colour television the signal amplitude is commonly standardized so as to make the signals R (red), G (green) and B (blue) of equal strength for white, the relation for white thus being:

$$R = G = B.$$

Moreover, it is usual to give the value of 1 to the signal values for the white with the greatest luminance occurring in the scene. Thus, for this white of maximum luminance

$$R = G = B = 1.$$

If this white is reproduced with a luminance L, the red primary colour contributes 0.30 L, the green 0.59 L, and the blue 0.11 L; the sum is, of course, exactly L. These contributions of luminance are therefore in this case the luminances of the primary colours corresponding to the signal values 1.

For the sake of convenience we shall now assume that the display device functions linearly, i.e. that for each primary colour the reproduced luminous flux is proportional to the signal strength R, G or B. For each value of R, G and B the luminance contributions will then be respectively 0.30 RL (red), 0.59 GL (green) and 0.11 BL (blue). We may then write for each colour that the luminance is equal to:

$$(0.30\ R + 0.59\ G + 0.11\ B)L.$$

The linear combination

$$H = 0.30\ R + 0.59\ G + 0.11\ B \quad . . \quad (10)$$

is called the luminance signal. For white with the maximum luminance ($R = G = B = 1$), the value $H = 1$. The luminance signal corresponds to the normal video signal of monochromatic television.

Still assuming the display device to function linearly, we can calculate as follows the signals R, G and B needed for producing a colour with coordinates X, Y, Z.

According to the normalizing of R, G and B mentioned above,

$R = 1$ corresponds to 0.298 L cd/m² of the primary red,
$G = 1$ to 0.588 L cd/m² of the primary green, and
$B = 1$ to 0.114 L cd/m² of the primary blue.

It now follows directly from (9a), (9b) and (9c) that for given R, G and B the coordinates X, Y and Z of the reproduced colour will satisfy:

$$X = 2.03\times0.298\ RL + 0.296\times0.588\ GL + 1.75\times0.114\ BL,$$
$$Y = 0.298\ RL + 0.588\ GL + 0.114\ BL,$$
$$Z = 0.113\times0.588\ GL + 9.75\times0.114\ BL.$$

For this we may write:

$$\frac{X}{L} = 0.604\ R + 0.174\ G + 0.200\ B,$$
$$\frac{Y}{L} = 0.298\ R + 0.588\ G + 0.114\ B,$$
$$\frac{Z}{L} = 0.000\ R + 0.0664\ G + 1.112\ B.$$

For R, G and B as functions of X, Y and Z we find from this:

$$\left.\begin{aligned} RL &= 1.92\ X - 0.535\ Y - 0.290\ Z,\\ GL &= -0.984\ X + 2.00\ Y - 0.0274\ Z,\\ BL &= 0.0588\ X - 0.119\ Y + 0.901\ Z. \end{aligned}\right\} \quad . . \quad (11)$$

In fact, normal display devices do *not* function linearly. This necessitates certain corrections to the above considerations (gamma correction), which we shall discuss at the end of this article.

Signals R, G and B at the transmitting end

We shall now say a few words about the generation of signals R, G and B at the camera end.

If $E(\lambda)$ is the spectral energy distribution of a given colour in the scene, we may write for the signals to be generated:

$$\left.\begin{aligned} R &= \int \bar{R}(\lambda)\ E(\lambda)\ d\lambda,\\ G &= \int \bar{G}(\lambda)\ E(\lambda)\ d\lambda,\\ B &= \int \bar{B}(\lambda)\ E(\lambda)\ d\lambda. \end{aligned}\right\} \quad . . . . . \quad (12)$$

$\bar{R}(\lambda)$, $\bar{G}(\lambda)$ and $\bar{B}(\lambda)$ are the total spectral sensitivities of the three channels. How must these be chosen in order to obtain exact colour reproduction? The necessary variation of $\bar{R}(\lambda)$, $\bar{G}(\lambda)$ and $\bar{B}(\lambda)$ as a function of λ can be calculated; the result (apart from the constant L/K) is shown in *fig. 9*. Exact colour reproduction is only possible for the colours inside the triangle RGB; only for these colours are none of the signals R, G and B negative.

From (11) and (6) it follows that for exact colour reproduction the signals R, G and B must satisfy:

$$\frac{L}{K}R = 1.92 \int \bar{X}(\lambda)\ E(\lambda)\ d\lambda - 0.535 \int \bar{Y}(\lambda)\ E(\lambda)\ d\lambda - 0.290 \int \bar{Z}(\lambda)\ E(\lambda)\ d\lambda,$$

$$\frac{L}{K}G = -0.984 \int \bar{X}(\lambda)\ E(\lambda)\ d\lambda + 2.00 \int \bar{Y}(\lambda)\ E(\lambda)\ d\lambda - 0.0274 \int \bar{Z}(\lambda)\ E(\lambda)\ d\lambda,$$

$$\frac{L}{K}B = 0.0588 \int \bar{X}(\lambda)\ E(\lambda)\ d\lambda - 0.119 \int \bar{Y}(\lambda)\ E(\lambda)\ d\lambda + 0.901 \int \bar{Z}(\lambda)\ E(\lambda)\ d\lambda.$$

By comparing these three expressions with (12) it follows directly that:

$$\frac{L}{K}\bar{R}(\lambda) = 1.92\,\bar{X}(\lambda) - 0.535\,\bar{Y}(\lambda) - 0.290\,\bar{Z}(\lambda),$$

$$\frac{L}{K}\bar{G}(\lambda) = -0.984\,\bar{X}(\lambda) + 2.00\,\bar{Y}(\lambda) - 0.0274\,\bar{Z}(\lambda),$$

$$\frac{L}{K}\bar{B}(\lambda) = 0.0588\,\bar{X}(\lambda) - 0.119\,\bar{Y}(\lambda) + 0.901\,\bar{Z}(\lambda).$$

$\bar{X}(\lambda)$, $\bar{Y}(\lambda)$ and $\bar{Z}(\lambda)$ are here known functions of λ, viz. the spectral distribution curves for the "spectrum of constant energy flux", which also occur in (6). See fig. 5.

It can be seen from fig. 9 that a negative sensitivity is required in certain wavebands. Since this requirement cannot easily be put into effect, and since its neglect results in only minor errors in the colour reproduction, the negative portions of the curves are normally left out of consideration.

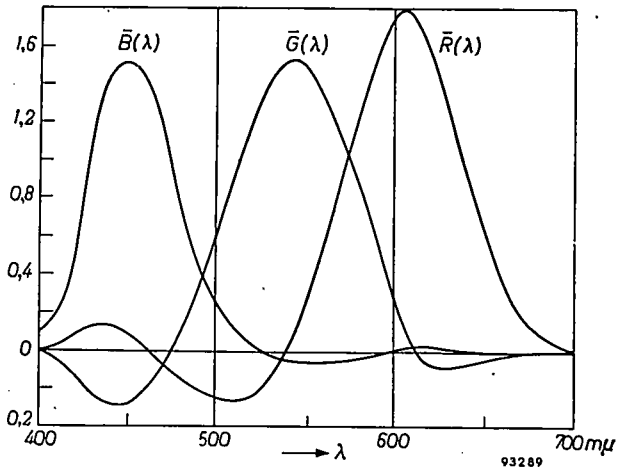


Fig. 9. For exact colour reproduction the quantities $\bar{R}(\lambda)$, $\bar{G}(\lambda)$ and $\bar{B}(\lambda)$ from (12) must, according to the theory, depend upon λ as shown.

In so far, however, as the spectral sensitivity of the camera tube in question, in combination with the selective operation of the dichroic mirror or mirrors, is not entirely in accordance with the curves in fig. 9, an attempt must be made to correct the error by introducing a filter in the path of the light. *Fig. 10* illustrates this for green light. The curves D_b and D_r in fig. 10*a* represent the transmission of the blue and red reflecting mirrors respectively. Multiplying the one ordinate by the other produces the curve D_bD_r, which shows how the light is transmitted which passes through both mirrors. It can be seen that a fair amount of blue is transmitted in addition to the required green. This will no longer happen if the light is also passed through the filter whose transmission characteristic D_f is given in fig. 10*b*. This figure also gives the transmission characteristic $D_bD_rD_f$ of the two mirrors and the filter together. The spectral sensitivity of the camera tube itself (a vidicon) is given by the curve V in fig. 10*c*, the total spectral sensitivity of the tube in combination with the two mirrors and the filter by the curve $D_bD_rD_fV$. The latter is

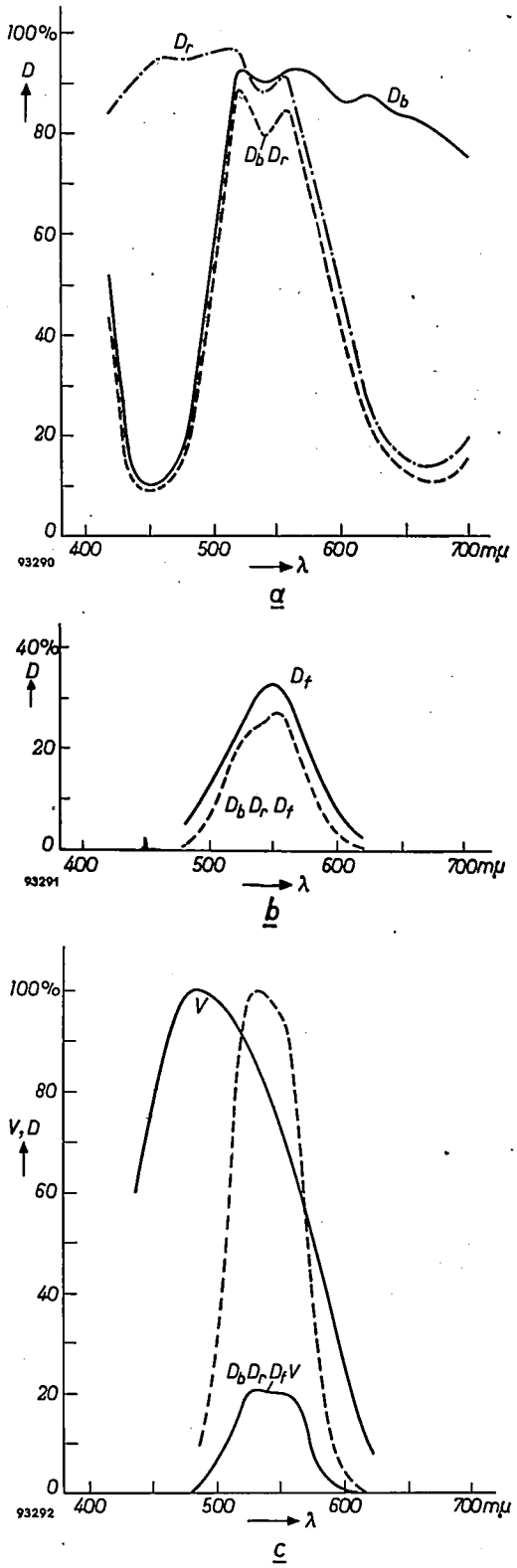


Fig. 10. *a*) Transmission characteristic of dichroic mirrors as a function of λ: D_r for a red reflecting mirror, D_b for a blue reflecting mirror. D_bD_r applies to light transmitted by both mirrors.
b) D_f transmission of a correction filter ("Kodak" 52). $D_bD_rD_f$ represents the transmission of both mirrors and the filter together.
c) V spectral sensitivity of the "green" camera tube (vidicon). $D_bD_rD_fV$ represents the total spectral sensitivity of camera tube with mirrors and filter. This has been re-drawn as a broken curve with its peak at 100%.

also drawn in such a way that the peak coincides with 100%. This curve represents an adequate approximation to the theoretical form of $\bar{G}(\lambda)$ in fig. 9.

Transmission systems

The conveyance of the three video signals in fig. 7 from the transmitter to the receiver by means of a carrier wave, on which they are modulated, requires in all a bandwidth three times as large as that needed for transmitting a monochromatic signal of equal definition. Obviously, ways and means have been sought to reduce this bandwidth without impairing the quality of the picture.

To begin with, we can see what the effect is of reducing the bandwidth of one or more of the primary signals. Slightly reducing the bandwidth of the green results immediately in a noticeable loss of definition. In the case of red, a slight reduction of bandwidth is much less disturbing. The bandwidth of the blue can be appreciably reduced before it adversely affects a normal picture. It appears that the bandwidth can be more limited the less the signal concerned contributes to the luminance ($H = 0.30\ R + 0.59\ G + 0.11\ B$).

An even smaller total bandwidth can be used if, instead of the signals R, G and B themselves, we transmit three mutually independent linear combinations of the signals, one of which is the luminance signal $H = 0.30\ R + 0.59\ G + 0.11\ B$; the two other linear combinations we shall call S_1 and S_2. At the receiving end the signals R, G and B are restored by forming the correct linear combinations of H, S_1 and S_2. The signal H determines the total luminance at the reproduction end, while S_1 and S_2 influence only the colour. The impression of sharpness of the resultant picture is determined primarily by the bandwidth of the luminance signal H. The bandwidth of the colour signals S_1 and S_2 can be strongly reduced without noticeably impairing the quality of the picture.

As already remarked, the luminance signal corresponds entirely to a normal monochromatic video signal. If the Gerber standard (the 625 line system of the Comité Consultatif International des Radiocommunications (C.C.I.R.)) is used with a video bandwidth of 5 Mc/s for the signal H, a bandwidth of, say, 1 Mc/s for S_1 and S_2 will be sufficient for producing a good colour picture. The total bandwidth will then be $5 + 1 + 1 = 7$ Mc/s, as against $5 + 5 + 5 = 15$ Mc/s if R, G and B are to be given the full bandwidth.

The system described, with the signals H, S_1 and S_2, is represented schematically in *fig. 11*. The colour signals at the input of the receiver *Rec* lack the components with frequencies above 1 Mc/s; to distinguish them from the others they are shown in square brackets: $[S_1]$ and $[S_2]$. Because of this difference, the primary colour signals R^*, G^* and B^*, which are formed in the receiver, are not the same as the original primary colour signals R, G and B. The components of R^*, G^* and B^* above 1 Mc/s do not differ from each other since they all originate from the luminance signal H. In America this is called a "mixed-highs" system. It follows from the foregoing that the resultant picture is a three-colour picture only for the components below 1 Mc/s; the components above 1 Mc/s, which provide the fine details, are added in black and white. In most cases the result is not perceptibly different from a picture composed entirely of three colours with all components up to 5 Mc/s.

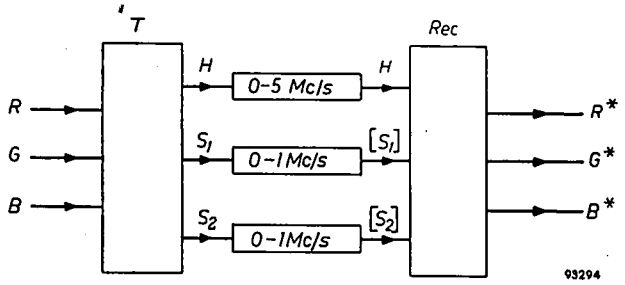


Fig. 11. Colour television system in which the transmitted signals consist of a luminance signal H and two colour signals, S_1 and S_2, which are linear combinations of the signals R, G and B. The signal H is radiated from the transmitter T to the receiver *Rec* with a bandwidth of 5 Mc/s, the signals S_1 and S_2 with a reduced bandwidth, e.g. of the order of 1 Mc/s.

If the following selection is made for S_1 and S_2:

$$\left.\begin{aligned} S_1 &= R - H, \\ S_2 &= B - H, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (13)$$

the signals R^* and B^* are obtained at the receiving end as the sum of the luminance signal H and the colour signal $[S_1]$ or $[S_2]$:

$$R^* = H + [S_1], \quad . \quad . \quad . \quad . \quad (14a)$$

$$B^* = H + [S_2]. \quad . \quad . \quad . \quad . \quad (14b)$$

The signal G^* is produced in a somewhat more intricate manner. It follows from (10) and (13), again indicating the limitation in bandwidth by square brackets, that:

$$0.30\ [R-H] + 0.59\ [G-H] + 0.11[B-H] = 0.$$

Hence

$$[G-H] = -\frac{0.30}{0.59}[S_1] - \frac{0.11}{0.59}[S_2], \quad . \quad (15)$$

and

$$G^* = H + [G-H]. \quad . \quad . \quad . \quad . \quad (16)$$

The four processes to be carried out at the receiving end in order to derive R^*, G^* and B^* from the available signals H, $[S_1]$ and $[S_2]$ are thus:

1) to add H and $[S_1]$: according to (14a), this produces R^*;
2) to add H and $[S_2]$: according to (14b), this produces B^*;
3) to combine certain fractions of $[S_1]$ and $[S_2]$: according to (15), this produces $[G-H]$;
4) to add H and $[G-H]$: according to (16), this produces G^*.

It may be added that (14a), (14b) and (16) can be written as follows:

$$\left.\begin{aligned} R^* &= H + [R] - [H] = [R] + H_h, \\ B^* &= H + [B] - [H] = [B] + H_h, \\ G^* &= H + [G] - [H] = [G] + H_h, \end{aligned}\right\} \quad (17)$$

in which $H_h = H - [H]$ represents the luminance signal components with high frequencies (the "mixed highs"). It is evident from (17) that R^*, G^* and B^* contain, as we have just seen, the same high-frequency components (H_h).

If the luminance signal H is modulated, like an ordinary monochromatic video signal, upon a carrier, and this is transmitted in a television channel together with a sound carrier, it will not differ in any way from a monochromatic transmission. Ordinary television receivers will reproduce such a transmission in normal black and white. This satisfies the requirement of "compatibility", which, in view of the large number of monochromatic receivers in many countries, is of great practical importance.

The question now remaining is, how must $[S_1]$ and $[S_2]$ be transmitted such that a colour television receiver will produce a coloured picture from them? Of course, $[S_1]$ and $[S_2]$ could be separately modulated on carriers outside a television channel. However, in the frequency allotment plan as now generally accepted (for Europe the 1952 Stockholm plan) there is no room for such a procedure. For this reason, means have been sought of accommodating the signals $[S_1]$ and $[S_2]$ in the channels already in use.

System with two subcarriers

The addition to a video signal of a frequency f which is an integral multiple of the line frequency f_l ($f = mf_l$) produces in the television picture a disturbing pattern of vertical strips. The reason for this is that the maxima created in each line by the interfering signal fall exactly below the corresponding maxima in the preceding line (*fig. 12a*), and in all subsequent scans the maxima recur at the same positions. However, if the added signal has a frequency that lies exactly halfway between two multiples of the line frequency: $f = (m + \frac{1}{2})f_l$, there will then be a shift of exactly 180° between the maxima and minima in successive lines (fig. 12*b*).

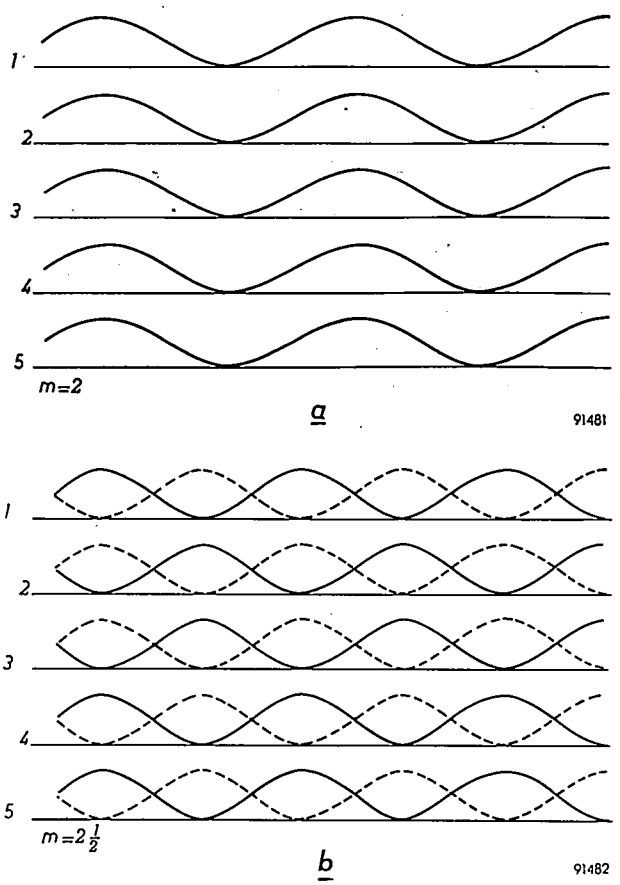


Fig. 12. Variation of luminance (as a result of an interfering signal with frequency f) of the lines 1, 3, 5, . . , 2, 4, . . . of a television scan, *a*) for f = twice the line frequency f_l, *b*) for $f = 2\frac{1}{2}f_l$. In (*a*) the maxima appear directly one above the other and occur at the same place in each successive complete scan. In (*b*) the maxima of the one line appear under the minima of the directly preceding line, and in successive complete scans the maxima and minima occur alternately (full line and broken line).

Moreover, since the total number of lines of a complete picture is uneven, the phase of the added signal is also shifted 180° at each successive scan of the same line (i.e. one picture interval later): at each point where a maximum occurs in the one complete scan, a minimum occurs in the next complete scan. Owing to the inertia of the eye, an added signal of this nature is much less troublesome than a signal with $f = mf_l$. For this reason a signal with $f = (m + \frac{1}{2})f_l$ may permissibly have a fairly large amplitude with hardly any adverse effect on the monochromatic picture, especially if f is not too low.

The above system can be used for adding to the

luminance signal a video-frequency carrier, modulated with e.g. the signal $[S_1]$, in such a way as to produce no interference in the monochromatic receiver. For this purpose, this subcarrier must have a frequency of $(m + \frac{1}{2})$ times the line frequency f_l. Although, strictly speaking, the above reasoning no longer holds for moving objects, it has been found in practice that even then the added signal may permissibly have a fairly large amplitude. The second colour signal $[S_2]$ might be modulated on a second subcarrier.

This was the manner of operation of one of the two systems which, in April 1955 and April 1956, were demonstrated in Eindhoven for the C.C.I.R. investigatory committee for colour television [7]).

N.T.S.C. system

Another possibility, which was also demonstrated in Eindhoven on the same occasion, has been developed more especially in America, where it is used for the present colour-television transmissions. In this "N.T.S.C. system" (so called after the National Television System Committee, which coordinates colour television work in the United States [8])) the signals $[S_1]$ and $[S_2]$ are again modulated on a subcarrier, but in this case the two subcarriers have the same frequency and are shifted 90° in phase with respect to each other:

$$[S_1] \cos \omega_a t + [S_2] \sin \omega_a t,$$

in which ω_a is the angular frequency of the subcarriers.

To separate $[S_1]$ and $[S_2]$ in the receiver it is now not possible to use conventional amplitude detectors, since these respond only to the total amplitude $\sqrt{[S_1]^2 + [S_2]^2}$. Separation can, however, be effected by means of synchronous detection. If we multiply (e.g. with the aid of a mixer tube) the total colour information signal

$$[S_1] \cos \omega_a t + [S_2] \sin \omega_a t$$

by $2 \cos \omega_a t$, we then obtain:

$$2[S_1] \cos^2 \omega_a t + 2[S_2] \sin \omega_a t \cos \omega_a t = \\ = [S_1] + [S_1] \cos 2\omega_a t + [S_2] \sin 2\omega_a t.$$

Provided the auxiliary carrier frequency $\omega_a/2\pi$ is high enough (greater than the bandwidth of $[S_1]$ and $[S_2]$), all that remains after passing through a suitable low-pass filter is the signal $[S_1]$. The signal $[S_2]$ is restored in an analogous way.

For this synchronous detection the auxiliary signals $\cos \omega_a t$ and $\sin \omega_a t$ are needed in the receiver. These are obtained from a wave train ("burst") sent out by the transmitter after each line synchronization pulse; the burst consists of a specific number (e.g. eight) cycles of the subcarrier frequency and has a specific phase, e.g. the cosine phase. In the receiver the burst can be used to synchronize an oscillator, from which the one auxiliary signal is obtained and, after a 90° phase shift, the other.

If, as under (13), S_1 is taken as $R-H$ and S_2 as $B-H$, each possibly multiplied by a constant factor

$$S_1 = \alpha(R-H) \text{ and } S_2 = \beta(B-H),$$

then the total colour information signal (see *fig. 13*) is:

$$S = \alpha(R-H) \cos \omega_a t + \beta(B-H) \sin \omega_a t. \quad (18)$$

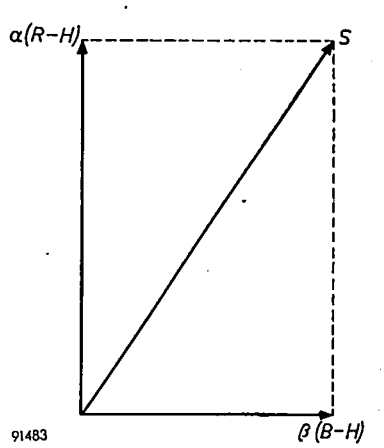


Fig. 13. The total colour information signal S according to the N.T.S.C. system (see eq. 18).

For white (and grey) this signal is zero. An increasing amplitude of the signal, the luminance signal remaining constant, corresponds to increasing colour saturation. A change of phase constitutes a change of the dominant wavelength of the colour.

In the American colour-television transmissions a small refinement is at present being employed, as described below.

The video spectrum according to (18) appears as drawn in *fig. 14a*. f_a is the frequency of the subcarrier wave on which $[S_1]$ and $[S_2]$ are modulated according to the normal double-sideband system. The latter is necessary in order to be able to separate the two colour signals by means of synchronous detection; with single-sideband modulation there would be complete cross-talk between these signals. The largest bandwidth that can be chosen for $[S_1]$ and $[S_2]$ in this system is thus the total video bandwidth (5 Mc/s in the Gerber system) minus the subcarrier frequency f_a.

7) J. Haantjes and K. Teer, Compatible colour-television, I. Two sub-carrier system, II. Comparison of two sub-carrier and N.T.S.C. systems, Wireless Engr. **33**, 3-9 and 39-46, 1956 (Nos. 1 and 2).

8) D. G. Fink, Color television standards; selected papers and records of the N.T.S.C., McGraw-Hill, New York 1955. Hazeltine Labs. Staff, Principles of color television, John Wiley & Sons, New York 1956.

The refinement referred to consists in giving one of the colour signals a larger bandwidth, by modulating the high-frequency components of the signal with only one sideband upon the subcarrier. At demodulation there is admittedly cross-talk between these components and the other colour signal, but this interference can be removed by means of a low-pass filter.

This system has been adopted in the present-day N.T.S.C. system. For the colour signal with enlarged bandwidth, one might choose between $\alpha(R-H)$ and $\beta(B-H)$, but it is possible to make a better choice. In the expression

$$S = S_1 \cos \omega_a t + S_2 \sin \omega_a t$$

the right-hand term is identical with:

$$(S_1 \cos \varphi - S_2 \sin \varphi) \cos (\omega_a t + \varphi) + \\ + (S_1 \sin \varphi + S_2 \cos \varphi) \sin (\omega_a t + \varphi)$$

for each value of the phase angle φ. The signal S can therefore also be formed by modulating the subcarrier wave in the phase $\cos (\omega_a t + \varphi)$ with the signal $(S_1 \cos \varphi - S_2 \sin \varphi)$ and in the phase $\sin (\omega_a t + \varphi)$ with the signal $(S_1 \sin \varphi + S_2 \cos \varphi)$, while still

$$S_1 = \alpha(R-H) \text{ and } S_2 = \beta(B-H).$$

Experiments have been made to ascertain for what value of φ an enlarged bandwidth of the signal to be modulated on $\cos (\omega_a t + \varphi)$ contributes most to improving the colour picture. With the values of α and β chosen in America the greatest improvement was obtained for $\varphi = 33°$. With this value the colour information signal is:

$$I \cos (\omega_a t + 33°) + Q \sin (\omega_a t + 33°),$$

in which

$$I = \alpha(R-H) \cos 33° - \beta(B-H) \sin 33° \quad . \quad . \quad (19)$$

thus gets a somewhat larger bandwidth than

$$Q = \alpha(R-H) \sin 33° + \beta(B-H) \cos 33°. \quad . \quad . \quad (20)$$

The upper sideband is moreover limited to the bandwidth of the signal Q. The result is illustrated in fig. 14*b*.

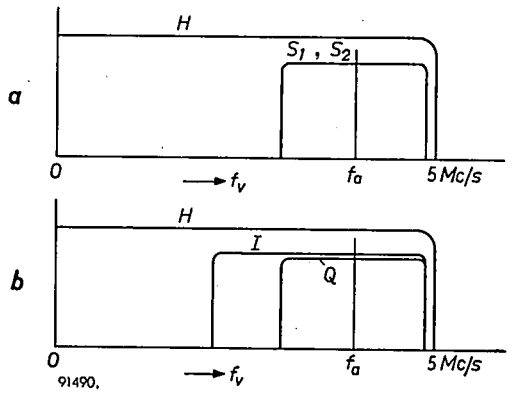


Fig. 14. *a*) Video-frequency spectrum in which the two colour signals, S_1 and S_2, have equal bandwidths. Each is modulated with two sidebands of limited width on a subcarrier with video frequency f_a; the two subcarriers are 90° out of phase with each other.
b) Video-frequency spectrum in which the two colour signals, I and Q, have different bandwidths (see (19) and (20)). In this case, I can have a wider lower sideband. This system is employed in America.

Variation in the amplitude of the signal I results in a change of colour along the line which in the colour triangle connects orange with blue-green (the line I in *fig. 15*). Information on colour variation along this line appears to be more important than information on variation along the green-purple line, which relates to variation of the signal Q (the line Q in fig. 15).

Gamma correction

As remarked, the relation between the luminous flux Φ and the control voltage V (measured from the cut-off point) is not linear in conventional picture tubes. By approximation this relation is a power function:

$$\Phi = c\, V^{\gamma}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

in which the exponent γ lies between 2 and 2.5. It is necessary to make a correction for this non-linearity (gamma correction), as otherwise impermissible errors would arise in the reproduction.

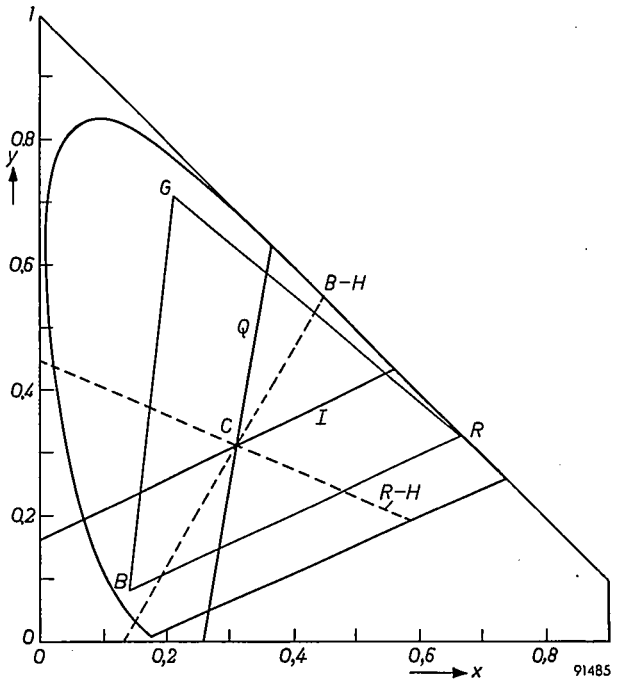


Fig. 15. Variation of signal I (eq. 19) causes a change of colour along line I in the colour triangle; variation of signal Q (eq. 20) causes a change along line Q. The colour points of the primary colours are R, G and B. The signals S_1 and S_2 correspond to the lines $R-H$ and $B-H$.

If every receiver were to be provided with such a correction, the price of the receivers would be considerably increased. For that reason, the gamma correction is made in the transmitter, by including a non-linear element (gamma corrector) in each of the three channels (red, green and blue) before the signals are made up into luminance and colour signals. If the expression (21) applies at the reproduction end for each of the three primary colours, the relation between the output voltage

V_o and the input voltage V_i of the gamma corrector must be:

$$V_o = c' V_i^{1/\gamma},$$

if, at least, the camera tubes supply a signal that is linearly dependent upon the luminance L_i on the photo-cathode. If this is not the case, the gamma corrector is made such that the relation between V_o and L_i is:

$$V_o = c'' L_i^{1/\gamma}.$$

If this is done in each colour channel, we obtain, instead of the signals R, G and B, the corrected signals

$$R' = R^{1/\gamma}, \quad G' = G^{1/\gamma}, \quad B' = B^{1/\gamma}.$$

The new luminance signal H' now becomes (eq. 10):

$$H' = 0.30\,R' + 0.59\,G' + 0.11\,B',$$

and the new colour signal S' (in the N.T.S.C. system) becomes

$$S' = \alpha(R'-H') \cos \omega_a t + \beta(B'-H') \sin \omega_a t.$$

In the receiver these signals are treated exactly as described above for the uncorrected signals, so that R', G' and B' are the signals which are fed to the reproduction device.

Gamma correction at the transmitting end has, it is true, certain disadvantages, which we shall not go into here; on the other hand it is indispensable for ensuring good reproduction without making the receivers all too complicated.

Summary. This introductory article on colour television begins by recapitulating some colorimetric concepts, such as additive colour mixing and the chromaticity diagram (colour triangle). The principle of colour television is then explained with reference to a system containing three camera tubes and three picture tubes, considered at this stage with separate transmission channels for the three primary colours (red, green and blue). The splitting of the incident light into the three primary colours at the transmitting end, and the combination of the three (projected) pictures at the receiving end, can be effected by means of dichroic mirrors.

As regards transmission the main problem is to limit the bandwidth. A discussion follows of a system with two subcarriers and of the N.T.S.C. system used in America. Finally, the author deals with the gamma correction needed in view of the non-linear relation between luminous flux and control voltage in picture tubes.

THE JUNCTION TRANSISTOR AS A NETWORK ELEMENT AT LOW FREQUENCIES

II. EQUIVALENT CIRCUITS AND DEPENDENCE OF THE *h* PARAMETERS ON OPERATING POINT

by J. P. BEIJERSBERGEN, M. BEUN and J. te WINKEL.

621.375.4

The primary purpose of this article is to make it clear how the h parameters of a junction transistor, as discussed in a previous article, depend on the biassing (D.C. operating point) and on the temperature. To this end recourse is made to what may be termed a "physical" equivalent circuit. At the same time the article deals with equivalent circuits in general, in order to show the place occupied by the physical type of equivalent circuit in relation to other types by which transistors can be represented.

It is common practice to represent a transistor by an electrical equivalent circuit. All such circuits are networks made up of linear circuit elements (resistances, self-inductances, capacitances, current and voltage sources, etc.), and thus they are only able to represent the transistor insofar as it behaves as a linear element. This being so, an equivalent circuit can only be made to represent the small-signal A.C. behaviour of a transistor, its behaviour with regard to D.C. being left entirely out of consideration. The same of course applies to equivalent circuits for electronic tubes.

The engineer concerned with the practical application of transistors makes use of an equivalent circuit with the aim of seeing what his circuits will do. In doing so, the properties of the transistor are no longer concealed behind a mysterious symbol but are represented in terms of elements of which the behaviour is known, which can be given numerical values and which can be compared qualitatively and quantitatively with other elements in the circuit. Various types of equivalent circuit can be used in this way. The endeavour will be to select the one that suits the rest of the circuit, and thus to obtain the highest possible degree of clarity and simplicity.

We shall begin by making one or two observations about such equivalent circuits as are suitable from this point of view. We shall then go on to discuss what we call a "physical" equivalent circuit, which we shall employ for an entirely different purpose, namely, to deduce how the *h* parameters of a transistor depend on the biassing (D.C. operating point) and on the temperature.

The present article forms a sequel to an earlier article [1]) on the transistor as a network element, which will hereinafter be referred to as I. Article I gave an account of the characteristic curves of transistors and the three possible ways of connecting up a transistor as a fourpole (viz. in the common-emitter, common-base and common-collector configurations) and contained a study of the transistor as a linear fourpole, in particular in the common-emitter configuration. Use was made of the *h* parameters, these being indexed with an e, b or c according as they related to the common-emitter, common-base or common-collector configurations. In the present article, as in I, we shall confine ourselves to frequencies so low that the static characteristics of the transistor represent its behaviour with sufficient accuracy. Accordingly, no self-inductances or capacitances will figure in the equivalent circuits.

Types of equivalent circuit suitable for practical use

A very great number of types is possible. For any linear fourpole (not necessarily a transistor) it is possible to draw an equivalent circuit based on a set of fourpole equations. If input voltage v_1 and output current i_2 are expressed in terms of input current i_1 and output voltage v_2, fourpole equations are obtained in the form involving the *h* parameters:

$$\left.\begin{aligned} v_1 &= h_{11}i_1 + h_{12}v_2 \,, \\ i_2 &= h_{21}i_1 + h_{22}v_2 \,. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

[1]) J. P. Beijersbergen, M. Beun and J. te Winkel, The junction transistor as a network element at low frequencies, I. Characteristics and *h* parameters, Philips tech. Rev. **19**, 15-27, 1957/58 (No. 1).

In *fig. 1* these equations are "translated" into an equivalent circuit. And, in fact, if expressions for v_1 and i_2 are written down for this circuit, the above fourpole equations will be recovered. Four of the six ways in which the fourpole equations can be written (a pair of independent variables can be

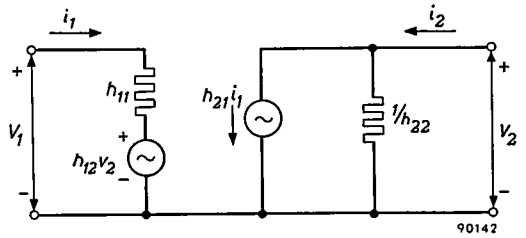


Fig. 1. An equivalent circuit for any linear fourpole can be drawn, given a set of fourpole equations for that fourpole. The above equivalent circuit is that corresponding to the equations involving the h parameters (equations (1)). The input loop contains a *voltage* generator of magnitude $h_{12}v_2$, the output loop a *current* generator of magnitude $h_{21}i_1$.

chosen from amongst v_1, v_2, i_1 and i_2 in six different ways) produce equivalent circuits [2]), all four circuits resembling fig. 1 in that each possesses two energy sources. This however does not exhaust the possibilities. An active linear fourpole is determined by four quantities; for this reason any equivalent circuit must contain four elements, at least one of which must be an energy source — otherwise the fourpole would not be active. But also any circuit having four elements can usually serve as equivalent circuit for an active linear fourpole. There therefore also exists a type of equivalent circuit having three resistances and only one voltage or current source.

The ability to judge which particular equivalent circuit — whether of the type with two energy sources or of the type with one — is best in a given case requires a good deal of experience, particularly if the case should be a complicated one. Often personal taste and instinct tip the scales. For those who are only incidentally concerned with the calculation of transistor circuits, the wisest course will often be to select the equivalent circuit that is most familiar, without worrying too much whether a different choice would make the calculations more elegant.

If a circuit with one energy source is decided on, the three resistances can be arranged in star or delta form. The first arrangement is usually drawn as a 'T' network and the second as a 'Π' network. The most useful types are the 'T' network of *fig. 2a* and the 'Π' network of fig. 2*b* [3]). The energy source in fig. 2*a* is a *voltage* generator proportional to the *input current* and located on the output side; that is, it is in series [4]) with r_2. That of fig. 2*b* is a *current* generator proportional to the input *voltage* and likewise located on the output side; here the energy source is in parallel [4]) with the resistance $1/y_2$.

Reasons for preferring the equivalent circuits of fig. 2

'T' and 'Π' networks other than those shown in fig. 2 can be drawn by making a different choice of energy source. There are various decisions to be made with regard to the energy source. Is it to be a current or a voltage generator? Is its magnitude to be made proportional to one of the currents or to one of the voltages involved in the fourpole? Which of the currents, or which of the voltages? Finally, where is the source to be located? Instead of systematically investigating the many possible combinations, we shall put forward arguments in favour of the equivalent circuits of figs. 2*a* and *b*.

We are here concerned with fourpoles that act as amplifiers, and we want to regard the internal feedback effect as a correction only. If there is no internal feedback, the source has to be placed on the output side and made proportional to one of the input quantities (input current or input voltage).

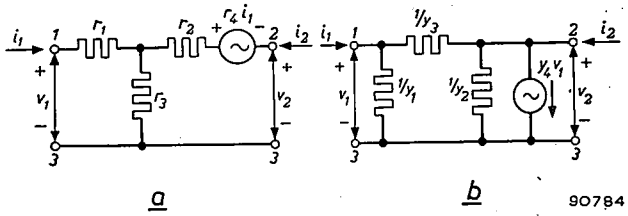


Fig. 2. Equivalent circuits for a linear active fourpole in the form of (*a*) a 'T' network and (*b*) a 'Π' network.

Any internal feedback can be allowed for by the presence in the circuit of a resistance r_3 (in fig. 2*a*) or $1/y_3$ (in fig. 2*b*). We leave the source in its place on the output side, however, and do not transfer it, for example, to the feedback path, although it would be quite possible to do so.

A 'T' network is particularly apt in cases where the elements of the external circuit are connected

[2]) Neither the set of equations expressing i_2 and v_2 in terms of i_1 and v_1 nor that expressing i_1 and v_1 in terms of i_2 and v_2 produces an equivalent circuit.

[3]) When the equivalent circuit is in the form of a 'Π' network, preference is given to the employment of admittances, denoted by y (the resistance values becoming $1/y$), in order that simple formulae may be obtained. The desirability of so doing will be apparent from formulae (2) and (3).

[4]) A voltage generator connected in *parallel* with a resistance would be equivalent to the voltage generator alone, for the latter — *by definition* — has zero internal resistance, which effectively short-circuits the resistance put in parallel. Similarly, a current generator (infinite internal resistance, by definition) in *series* with a resistance would be equivalent to the current generator alone.

in series to the terminal pairs of the fourpole. A voltage generator is generally preferred as source in a 'T' network, because it suits the character of a series arrangement [4]). When a 'T' network is chosen as equivalent circuit, the voltages can be immediately expressed in terms of the currents, providing that the magnitude of the source has been made proportional to a *current*, as has in fact been done in fig. 2*a*. (The reader may check that the expressions for v_1 and v_2 in terms of the currents alone can be written down simply by inspection of fig. 2*a*.)

A 'Π' network is highly suitable when the elements of the external circuit are connected in parallel across the terminal pairs of the fourpole. A current generator in parallel with $1/y_2$ suits this arrangement [4]). In a 'Π' network the currents can be immediately expressed in terms of the voltages, providing the magnitude of the source is made proportional to a *voltage*, as is done in fig. 2*b*. (The convenience of this procedure is demonstrated in the next section, equations (2) and (3).)

If desired, however, departures may be made from the above rules. Thus, when the fourpole represents a transistor in the common-emitter or common-base configurations, we frequently find a 'T' network being used which has a current generator in parallel with r_2 instead of a voltage generator in series with that resistance [5]) (as in fig. 2*a*). If we write down the magnitude of this current generator in terms of the h parameters of the fourpole, we obtain an expression $i_1(h_{21}+h_{12})/(1-h_{12})$ which, for a transistor in the aforementioned two configurations, is almost equal to $h_{21}i_1$, that is, to the current amplification factor multiplied by the input current. This has the advantage that the current amplification factor h_{21}, which frequently plays an important role in the said configurations (see for example I, p. 22), now figures in the equivalent circuit as an explicit quantity. However, to say that the 'T' network having a current source in parallel with r_2 is *preferable* because it is closely connected with the physical action of a transistor (as is sometimes argued) is not wholly correct.

Evaluation of the elements of an equivalent circuit

The equivalent circuit for a fourpole having been settled on, the numerical values of its elements must be known if it is to be of any practical use. In general it will not be possible to measure the values of the elements directly. For the elements of fig. 2*a* to be directly measured, it would be necessary for the nexus of the 'T' to be accessible, and this is certainly not the case when the fourpole is, for example, a transistor, for the nexus in the equivalent circuit does not correspond to an actual point in the transistor. Consequently formulae are required that express the elements of the selected equivalent circuit in terms of quantities that are accessible to direct measurement; preference will of course be given to those quantities that are most easily measured. Where the fourpole is a transistor, such quantities would be the h^e parameters (see I, p. 25) or alternatively the h^b parameters, which are equally easy to determine by measurement.

No fundamental difficulty is involved in the derivation of the formulae. We shall go through two examples to illustrate the general procedure.

Suppose that we want to express the elements of the 'Π' network in fig. 2*b* in terms of the h parameters of the fourpole of which it is the equivalent circuit. In fig. 2*b* we have:

$$\left.\begin{aligned} i_1 &= v_1y_1 + (v_1-v_2)y_3 = (y_1+y_3)v_1 - y_3v_2, \\ i_2 &= v_2y_2 + y_4v_1 + (v_2-v_1)y_3 = \\ &= (y_4-y_3)v_1 + (y_2+y_3)v_2. \end{aligned}\right\} \quad (2)$$

From (1) we now derive new expressions for i_1 and i_2 in terms of v_1 and v_2 and in which the coefficients are functions of the h parameters. By equating the coefficients in these expressions with the corresponding coefficients of (2), we obtain four further equations, the solution of which gives us the values of the unknown elements. The results, which are set out in the first column of *Table I*, are valid for any fourpole. If the fourpole represents a transistor, certain simplifications may be made (see Table I).

As an example of a more complicated case, let us suppose that we want to use the equivalent circuit of fig. 2*b* for a transistor in the common-base configuration (see *fig. 3a*, the elements of which have accordingly been given an index b), whereas only

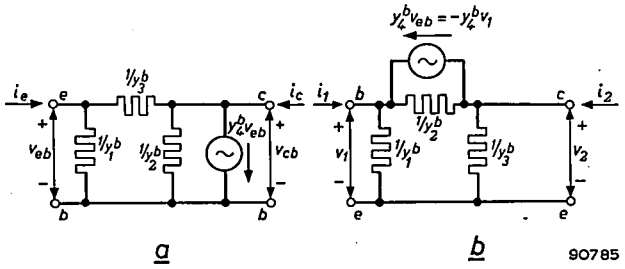


Fig. 3. *a*) The 'Π' network of fig. 2*b* taken as the equivalent circuit for a transistor, the base terminal *b* being the common terminal. *e*: emitter terminal. *c*: collector terminal.
b) The same network with the emitter terminal as the common terminal.

the h^e parameters are known to us. We therefore have to find expressions for the elements of fig. 3*a* that involve the h^e parameters. (The formulae in the first column of Table I cannot be used for this purpose, of course, because they would simply give us the elements expressed in terms of h^b parameters.)

[5]) A voltage generator of magnitude E in series with a resistance R is equivalent to a current generator of magnitude E/R in parallel with the same resistance R. This can easily be confirmed by comparing open-circuit voltage and short-circuit current in the two cases.

Table I. *First column:* The elements of a 'Π' network serving as equivalent circuit for a fourpole (see fig. 2*b*), expressed in terms of the h parameters of that fourpole.
Second column: The same, except that here the fourpole represented is a transistor in the common-emitter configuration (all quantities are accordingly given an index e above and to the right). Some simplifications have been rendered possible by the fact that $h^e_{12} \ll 1$.
Third and fourth columns: The elements of the 'Π' network expressed in terms of the h^e parameters of a transistor, in the cases where the network serves as equivalent circuit for the transistor in the common-base and common-collector configurations respectively.
Expressions for the slope S appear at the foot of each column.

Any linear fourpole	Transistor configuration		
	Common-emitter	Common-base	Common-collector
$y_1 = \frac{1-h_{12}}{h_{11}}$	$y_1^e = \frac{1-h_{12}^e}{h_{11}^e} \approx \frac{1}{h_{11}^e}$	$y_1^b = \frac{h_{21}^e+1}{h_{11}^e}$	$y_1^c = \frac{h_{12}^e}{h_{11}^e}$
$y_2 = h_{22} - \frac{h_{12}}{h_{11}}(h_{21}+1)$	$y_2^e = h_{22}^e - \frac{h_{12}^e}{h_{11}^e}(h_{21}^e+1)$	$y_2^b = \frac{h_{12}^e}{h_{11}^e}$	$y_2^c = h_{22}^e + \frac{h_{21}^e}{h_{11}^e}(1-h_{12}^e) \approx \frac{h_{21}^e}{h_{11}^e}$
$y_3 = \frac{h_{12}}{h_{11}}$	$y_3^e = \frac{h_{12}^e}{h_{11}^e}$	$y_3^b = h_{22}^e - \frac{h_{12}^e}{h_{11}^e}(h_{21}^e+1)$	$y_3^c = \frac{1-h_{12}^e}{h_{11}^e} \approx \frac{1}{h_{11}^e}$
$y_4 = \frac{h_{12}+h_{21}}{h_{11}}$	$y_4^e = \frac{h_{12}^e+h_{21}^e}{h_{11}^e} \approx \frac{h_{21}^e}{h_{11}^e}$	$y_4^b = -\frac{h_{12}^e+h_{21}^e}{h_{11}^e} \approx -\frac{h_{21}^e}{h_{11}^e}$	$y_4^c = -\frac{h_{12}^e+h_{21}^e}{h_{11}^e} \approx -\frac{h_{21}^e}{h_{11}^e}$
$S = y_4 - y_3$	$S = y_4^e - y_3^e \approx y_4^e$	$S = y_4^b - y_3^b \approx y_4^b$	$S = y_4^c - y_3^c \approx y_4^c$

Before the desired formulae can be derived one point must be noted: although in accordance with usual practice the equivalent circuits under discussion have been drawn as fourpoles, they actually represent threepoles; the terminals marked *3* in fig. 2*a* and *b* are shared by one connection. The circuit of fig. 3*a* therefore represents a transistor (which is, of course, a threepole) irrespective of the way in which it is connected in the circuit. In order to carry out the calculation we have in view, we shall redraw the circuit of fig. 3*a*, this time giving the common terminal to the emitter instead of the base (fig. 3*b*). We now write expressions for i_1 and i_2 in terms of v_1 and v_2, making use of the fact that in fig. 3*b*, $v_1 = v_{be} = -v_{eb}$. We find that:

$$\left.\begin{aligned} i_1 &= v_1 y_1{}^b + (v_1 - v_2) y_2{}^b + y_4{}^b v_1 = \\ &= (y_1{}^b + y_2{}^b + y_4{}^b) v_1 - y_2{}^b v_2, \\ i_2 &= v_2 y_3{}^b - y_4{}^b v_1 + (v_2 - v_1) y_2{}^b = \\ &= -(y_2{}^b + y_4{}^b) v_1 + (y_2{}^b + y_3{}^b) v_2. \end{aligned}\right\} \quad (3)$$

We again equate the coefficients of v_1 and v_2 with those we obtained from (1) on writing expressions for i_1 and i_2 in terms of v_1 and v_2 with the aid of the h^e parameters. We use the four equations that result to obtain expressions involving the h^e parameters for the unknown elements (y_1^b etc. in third column of Table I).

If fig. 2*b* is to represent a transistor in the common-collector configuration, we can in a similar way find expressions involving h^e parameters for the elements in the equivalent circuit (y_1^c etc. in fourth column of Table I). The numerical values given by the formulae of Table I for a Philips OC.71 transistor biassed to a normal D.C. operating point appear in *Table II*. It should be stressed that the Table I formulae only relate to certain arbitrarily selected examples; they are not of greater importance than other formulae that may be derived, as the need arises, in a similar way.

Table II. Numerical values given by the formulae of Table I for a Philips OC 71 transistor. At a D.C. operating point of $I_e = -3$ mA and $V_{ce} = -2$ V, this transistor has $h_{11}^e = 800\ \Omega$, $h_{12}^e = 5.4 \times 10^{-4}$, $h_{21}^e = 47$ and $h_{22}^e = 88 \times 10^{-6}\ \Omega^{-1}$. S, the slope, is found to be independent of the configuration.

	Configuration		
	Common-emitter	Common-base	Common-collector
y_1	$1.25 \times 10^{-3}\ \Omega^{-1}$	$6 \times 10^{-2}\ \Omega^{-1}$	$6.7 \times 10^{-7}\ \Omega^{-1}$
y_2	$4.8 \times 10^{-5}\ \Omega^{-1}$	$6.7 \times 10^{-7}\ \Omega^{-1}$	$6 \times 10^{-2}\ \Omega^{-1}$
y_3	$6.7 \times 10^{-7}\ \Omega^{-1}$	$4.8 \times 10^{-5}\ \Omega^{-1}$	$1.25 \times 10^{-3}\ \Omega^{-1}$
y_4	$5.9 \times 10^{-2}\ \Omega^{-1}$	$-5.9 \times 10^{-2}\ \Omega^{-1}$	$-5.9 \times 10^{-2}\ \Omega^{-1}$
S	59 mA/V	59 mA/V	59 mA/V

Slope of a transistor in the three configurations

Once the elements of a certain equivalent circuit have been evaluated for all three transistor configurations (as has just been done, with the results shown in the last three columns of Table I), the results can be used for comparing the properties of the transistor in each of the configurations with its properties in the others. By way of example, we

shall demonstrate that the slope S of the transistor, disregarding its sign, has almost (but not exactly) the same value in all three configurations. As may be found from the second equation of (2) by putting $v_2 = 0$, it is true of any fourpole that $S = (i_2/v_1)_{v_2=0} = y_4 - y_3$. Table I shows that y_4 has exactly the same value in all three configurations (if the sign be disregarded). We further find from Table II that y_3 is about 2% of y_4 in the common-collector configuration and even less in the two other configurations. Therefore S is practically equal to y_4, and independent of the configuration. We may further infer that where fig. 2*b* represents a transistor, the magnitude of the current generator may be taken to be Sv_1.

Physical equivalent circuit for a junction transistor

In general it is not possible to establish direct relationships between the elements figuring in an arbitrary equivalent circuit and the physical factors that govern the functioning of a transistor, such as the properties of the semiconductor material, the geometry of the transistor, its D.C. operating point and its temperature. However, for the junction transistor there does exist an equivalent circuit which allows these relationships to be readily established. In this, the "physical" type of equivalent circuit, the effects of the factors named above are brought to light. This type of equivalent circuit is not sufficiently tractable for circuit analysis; nevertheless, it is possible to express the elements of tractable equivalent circuits in terms of those of the physical one, and this allows the effect of physical factors on elements in tractable equivalent circuits to be investigated. In drawing the physical equivalent circuit we shall make use of the results obtained in an article [6]) which appeared earlier in this Review and in which the physical basis of transistor action was treated.

Let us consider a transistor in the common-base configuration connected up as a fourpole. We shall regard its input voltage V_{eb} and output voltage V_{cb} as being identical to the voltage across the emitter-base P-N junction and the collector-base junction respectively (see *fig. 4*; the upper-case letters V and I denote direct voltages and currents; lower-case letters v and i are used for the alternating voltages and currents that are superimposed upon them). Input and output voltages are not *strictly* the same as those across the junctions, for the base always has a certain internal resistance; we shall return to this point later on. The following formula for the input resistance for alternating current when the output is short-cuircuited, i.e. when V_{cb} is kept constant, was derived on p. 244 of article [6]):

$$R_0 = \frac{kT}{qI_e} \quad \ldots\ldots \quad (4)$$

(k is Boltzmann's constant, T the absolute temperature, q the absolute value of the electronic charge).

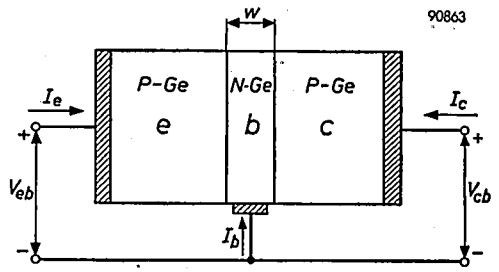


Fig. 4. Schematic diagram of a P-N-P junction transistor. Between the emitter e and the collector c, both of which are composed of P-germanium, is the thin base layer b, composed of N-germanium and having a thickness w. Each of the three regions has its own connecting wire.

The short-circuited output current is $-\alpha i_1$, the positive quantity α being the "physical" current amplification factor [7]). Hence the situation when the *output* is short-circuited for an alternating current is represented by the equivalent circuit of *fig. 5a*.

The situation in which the *input* is short-circuited for alternating current and in which there is an alternating voltage superimposed on the direct voltage at the output also lends itself to physical investigation (see the appendix to this article). It is then found that the ratio between the alternating voltage and the alternating current at the output — this being the output resistance with input short-circuited — is given by μR_0, where

$$\mu = \frac{2qw}{kT}\sqrt{\frac{qn_D|V_{cb}|}{2\varepsilon}} \quad \ldots\ldots \quad (5)$$

(w is the thickness of the base (see fig. 4), n_D is the excess of donor over acceptor concentrations in the base [8]) and ε is the dielectric constant of the

[6]) F. H. Stieltjes and L. J. Tummers, Simple theory of the junction transistor, Philips tech. Rev. **17**, 233-246, 1955/56.

[7]) See [6]), p. 239-240, and particularly formula (13), which expresses the effect of various physical quantities on α. Apart from the quantities appearing in this formula, α is also affected by the current density. This matter was discussed by F. H. Stieltjes and L. J. Tummers in Philips tech. Rev. **18**, 61-68, 1956/57 (No. 2).

[8]) Regarding the terms "donor" and "acceptor", see for example J. C. van Vessem, The theory and construction of germanium diodes, Philips tech. Rev. **16**, 213-224, 1954/55, particularly pp. 214 and 215.

transistor material). This situation is represented by the equivalent circuit of fig. 5*b*. On the input side a short-circuit current of value $-\beta i_2$ is flowing, the symbol β denoting the base efficiency as discussed on a previous occasion [9]). The general case is given by fig. 5*c*. It should be noted that in this combined equivalent circuit the current generators are expressed in terms, not of the input and output currents i_1 and i_2, but of the currents flowing through the resistances, indicated by symbols with primes to distinguish them from i_1 and i_2. It may easily be seen that fig. 5*c* transforms into fig. 5*a* when the output is short-circuited, and into fig. 5*b* when the input is short-circuited.

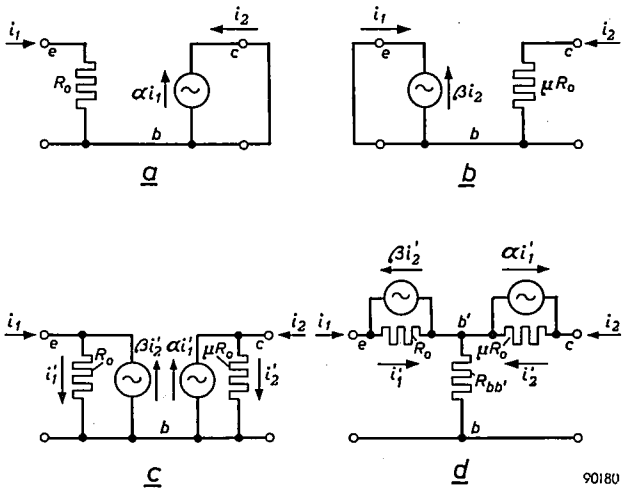


Fig. 5. *a*) Equivalent circuit for a junction transistor having zero internal base resistance, the transistor being in the common-base configuration and having its *output short-circuited;* the circuit is composed of elements whose values are easily deducible from the theory of the transistor. α is what we may call the "physical" current amplification factor.
b) Equivalent circuit as in (*a*), but appropriate to a transistor with its *input* short-circuited. For the value of μ see formula (5). β is the base efficiency (see Appendix and [6])).
c) Equivalent circuit produced by combining (*a*) and (*b*). The current generators are here made proportional to the currents i_1' and i_2' through the resistances, the prime serving to distinguish them from the input and output currents.
d) Equivalent circuit (*c*) completed by the insertion of $R_{bb'}$, the internal base resistance.

The transistors most often used in practice (including all Philips transistors) are junction transistors of the alloy type. The structure is shown diagrammatically in *fig. 6*. A transistor of this type consists of a single crystal of *N*-germanium (if it is a *P-N-P* transistor) in the form of a thin slab, portions on either side of which are turned into *P*-germanium by alloying with a material that acts as an acceptor, e.g. indium. Excesses of the acceptor material protrude on both sides and are used for affixing the emitter and collector connections. The base current enters via the base contact, which is soldered on to the edge of the tiny crystal slab. On its way to the true base region between emitter and collector, the base current encounters a certain amount of resistance, this forming the greater part of what is termed the internal base resistance. A smaller portion of the internal base resistance is attributable to the fact that the base current has to flow in a direction parallel to the *P-N* junctions in order to distribute itself over the entire cross-section of the effective part of the transistor. It is found that the internal base resistance may be satisfactorily accounted for by inserting a constant resistance $R_{bb'}$ in series with the base point *b* in fig. 5*c*. This point, which is marked *b'* in the complete equivalent circuit drawn in fig. 5*d*, is often called the "internal base"; it cannot be referred to a definite place in the transistor crystal. Obviously, $R_{bb'}$ is the greater the lower the conductivity of the base material.

Dependence of h^e parameters on D.C. operating point and on temperature

In order to investigate how the h^e parameters — and hence the elements of the equivalent circuit of fig. 1 — depend on the D.C. operating point and on temperature, we shall write down these parameters in terms of elements in the physical equivalent circuit. The procedure is similar to that by which, earlier on, we found expressions for the elements of an equivalent circuit in terms of h^e parameters. The results appear in *Table III*.

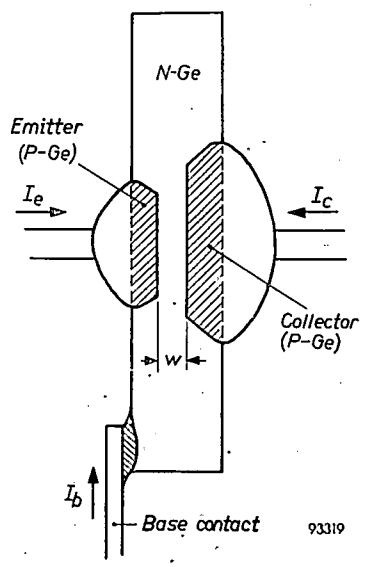


Fig. 6. Diagram showing schematically the construction of a *P-N-P* junction transistor of the alloy type. Most transistors, including all Philips transistors, are of this type.

[9]) See [6]), p. 240, formula (12), as regards the effect of various quantities on β. β is also affected by current density; in this connection see the article cited at the end of footnote [7]).

Table III. The h^e parameters of a transistor expressed in terms of the elements of the "physical" equivalent circuit of fig. 5*d*. The approximations given on the right are possible by virtue of the fact that μ is much greater and $(1-\beta)$ much less than unity, and that $(1-\alpha)$, though much less than unity, is greater than $(1-\beta)$.

$$h^e_{11} = R_{bb'} + \frac{\mu R_0}{1-\beta+(1-\alpha)\mu} \approx R_{bb'} + \frac{R_0}{1-\alpha}.$$

$$h^e_{12} = \frac{1-\beta}{1-\beta+(1-\alpha)\mu} \approx \frac{1-\beta}{\mu(1-\alpha)}.$$

$$h^e_{21} = \frac{\mu\alpha-1}{1-\beta+(1-\alpha)\mu} \approx \frac{\alpha}{1-\alpha}.$$

$$h^e_{22} = \frac{1-\alpha\beta}{1-\beta+(1-\alpha)\mu}\cdot\frac{1}{R_0} \approx \frac{1-\alpha\beta}{1-\alpha}\,\frac{1}{\mu R_0}.$$

At D.C. operating points and temperatures such as are normal in transistor work, it is usually possible to treat α, β, and $R_{bb'}$ as constants. This simplifies matters considerably, for it means that we have only the effect of R_0 and μ on the h^e parameters to take into account. The D.C. operating point is given by I_c, the output biassing current, and V_{ce}, the output biassing voltage (see article I, p. 20). Now, if signs be disregarded, $I_c \approx I_e$ and $V_{ce} \approx V_{cb}$ (see article I, p. 16); it follows from (4), therefore, that R_0 is inversely proportional to I_c and directly proportional to T, and from (5) that μ is directly proportional to $\sqrt{|V_{ce}|}$ and inversely proportional to T. We then see (Table III) that h^e_{21} is independent of both operating point and temperature; that h^e_{11} is independent of the biassing voltage but that it contains, besides a constant term ($R_{bb'}$), a term containing R_0 which is inversely proportional to the biassing current and directly proportional to the temperature; that h^e_{12} contains $1/\mu$ and is therefore inversely proportional to the root of the biassing voltage, directly proportional to the temperature, but independent of the biassing current (apart from any variation in α); and that h^e_{22} contains $1/\mu R_0$ and is therefore inversely proportional to the root of the biassing voltage, directly proportional to the biassing current, but independent of temperature. Of these conclusions, the most important in practice is that concerned with h^e_{11} (input resistance with output short-circuited), namely that h^e_{11} decreases with increasing collector biassing current. Its importance will be evident from an article on a transistor hearing-aid that will shortly be published in this Review.

A further point deserving of attention is the straightforward connection between h^e_{21}, the current amplification factor in the common-emitter configuration (in the literature h^e_{21} is frequently given the symbol α' or α_{cb}) and α, the "physical" current amplification factor. Since α is always a little less than unity, h^e_{21} is always considerably greater than unity — as was pointed out in article [6]).

Relative variations in absolute temperature are only slight in practice. It therefore follows from the foregoing that the *direct* effect of temperature on the A.C. behaviour of a transistor is also slight. However, depending on the D.C. properties of the external circuit, it is possible for temperature to exercise a considerable effect on the D.C. operating point and hence, indirectly, on the behaviour of the transistor with regard to alternating current. This and other questions connected with the effect of temperature on the operating point will form the subjects of a third and final article of this series on the junction transistor as network element at low frequencies.

Appendix: Allowance for the Early effect in the physical equivalent circuit

The expression for the resistance R_0 (formula (4)), and the strength of the current generator in fig. 5*a* (viz. αi_1) was taken over from the theory put forward in article [6]). On the other hand, the value for the resistance (μR_0) and the strength of the current generator (βi_2) in fig. 5*b* still require justification from the physical viewpoint. For this purpose we may refer to *fig. 7*,

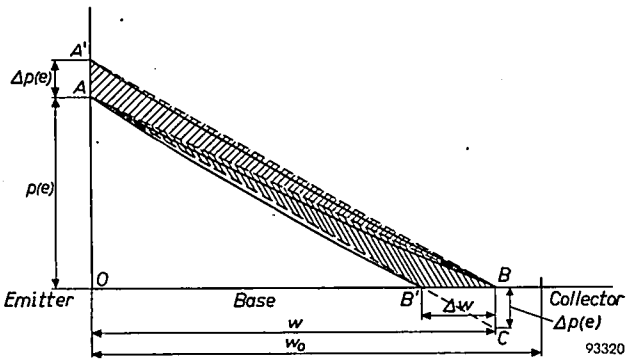


Fig. 7. The unbroken curve AB represents the hole concentration across the base of a *P-N-P* transistor. If the absolute value of V_{cb} is increased and V_{cb} kept constant, the new concentration curve is represented by AB'. $A'B$ represents the concentration curve when, on the other hand, V_{cb} is kept constant and V_{cb} is increased. The hatched triangles represent the associated changes in the base loss; in the first case the loss decreases, in the second it increases. The straight dashed lines are approximations to the actual hole concentration curves (full lines).

which shows the concentration curve AB of the minority carriers (holes) in the base of a *P-N-P* transistor (concentration curves were dealt with in article [6])). On the emitter side the concentration has a high value $p(e)$ determined by V_{eb}, the biassing voltage in the forward direction across the emitter-base *P-N* junction. At the collector the concentration is kept effectively at zero by a relatively high voltage in the inverse direction acting across the collector-base *P-N* junction. The hole current from emitter to collector is a diffusion current and is proportional at any point in its path to the concentration gradient. If V_{eb} is kept constant, $p(e)$ is constant; if at the same time V_{cb} is altered, then the concentration gradient in the base — and consequently the hole current across the base — alters too, owing to horizontal displacement of point B in fig. 7 (this is the so-called Early effect; see article [6]),

p. 245). The distance between the point B, where the hole concentration is zero, and the N-P junction between base and collector is, for the usual values of V_{cb}, given by $b\sqrt{-V_{cb}}$, the constant b being equal to [10]):

$$b = \sqrt{\frac{2\varepsilon}{qn_D}}. \quad \ldots\ldots\ldots \quad (6)$$

If, therefore, w_0 is the thickness of the base as measured to the N-P junction between base and collector, and if w is the *effective* thickness of the base, as measured to point B, then:

$$w = w_0 - b\sqrt{-V_{cb}}. \quad \ldots\ldots\ldots \quad (7)$$

Now, for the common-base configuration which we are considering, the output resistance with input short-circuited is:

$$R_u = 1 \Big/ \left(\frac{\partial I_c}{\partial V_{cb}}\right)_{V_{cb}}.$$

The problem is, therefore, one of evaluating $(\partial I_c/\partial V_{cb})_{V_{cb}}$. Since the electron current at the collector is constant (see article [6]), p. 239, column 2), we need only take the hole current into account. This current (I_c^+) is proportional to the hole concentration gradient at the collector. The gradient is given approximately by $p(e)/w$, and hence the hole current by

$$I_c^+ \approx -q\,D_p\,\frac{p(e)}{w} \quad \ldots\ldots\ldots \quad (8)$$

(D_p is the diffusion constant for holes). Of the holes in transit from emitter to collector, a fraction is continually being lost by recombination with electrons, so that the fall-off in concentration through the base is not quite linear. Hence at the emitter the gradient is somewhat steeper and at the collector it is somewhat less steep than $p(e)/w$ (see fig. 7). For the calculation of R_u, however, (8) gives I_c^+ with sufficient precision. On combining this expression with (7) and differentiating with respect to V_{cb}, we find that:

$$R_u = \frac{2w\sqrt{|V_{cb}|}}{b(-I_c^+)}.$$

But since $-I_c^+ \approx I_e$,

$$R_u = \frac{2w\sqrt{|V_{cb}|}}{bI_e},$$

which can be written $R_u = \mu R_0$, μ and R_0 being as defined in (5) and (4).

With the aid of fig. 7 we can, further, evaluate the current generator of fig. 5*b*. The need for a current generator in that equivalent circuit follows from the fact that, if the input voltage V_{eb} is kept constant and the output voltage V_{cb} altered, there will be changes in the input current I_e (feedback) as well as in the output current I_c. The reason is that in these circumstances there are changes in the hole current crossing the base, as already stated in the previous column. If there were no loss of holes in the base, both collector and emitter current (sign being disregarded) would change by the same amount as the hole current across the base. The magnitude of the current generator would then have to be i_2 exactly. In the event, the loss of holes in the base makes a correction necessary. In order to determine what the correction should be, we may recall that the area under the concentration curve AB in fig. 7, which we may take to be that of triangle ABO, is a measure of the said loss (see article [6]), p. 240). If B shifts to B' (owing to a change in V_{cb}) without A shifting (because V_{cb} is being kept constant), then the hole current undergoes an overall increase i^+ that is proportional to the increase in the average gradient. But at the emitter the increase is somewhat *less* than the increase at the collector, for the base loss has *decreased* by the area of triangle $AB'B$. It was observed above (first column of this page) that on the collector side the electron current is constant; hence the change undergone by the hole current on the collector side constitutes the whole of the current change that has taken place there. The same applies to the emitter side since V_{eb} is kept constant. Thus we see that the change in the total emitter current is less than that in the total collector current; hence, in the current generator of fig. 5*b*, the coefficient of i_2 must be less than unity. That the coefficient is exactly equal to β, the base efficiency, will be clear from the following considerations. If we keep B in place (by keeping V_{cb} constant) and shift A to a point A' such that $A'B$ is parallel to AB' (this shift being brought about by altering V_{eb} by a suitable amount), the average gradient of the hole concentration will have increased by the same amount as in the previous case, and hence the hole current will have increased overall by the same amount i^+. But now the increase at the emitter is somewhat *greater* than that at the collector, for the base loss has *increased* by the area of triangle $AA'B$. The areas of triangles $AA'B$ and $AB'B$ are equal to a first approximation, since their difference (this being the area of triangle $B'BC$) is an order of magnitude smaller. The change in the loss, $\Delta\{I_e^+ - (-I_c^+)\}$, in relation to the change in the hole current, ΔI_c^+, is therefore the same in the two cases. In the second case this ratio is equal to $1-\beta$ (this in fact being the definition of β, see [6]), eq. 12); this, then, is likewise the value of the ratio in the first case. Therefore the magnitude allotted to the current generator in fig. 5*b* must be βi_2.

[10]) W. Shockley, Bell Syst. tech. J. **28**, 450, 1949. It may be noted that V_{cb} is a negative quantity, so that $-V_{cb}$ is positive.

Summary. A transistor, considered as a linear network element, may be represented by many different types of equivalent circuit. Normally an equivalent circuit will be used to get an idea of how the transistor behaves in a proposed circuit. In such a case there is no point in trying to find a "physical" equivalent circuit; the type selected should be one suited to the circuit of which the transistor forms part. Mention is made of the types most commonly used for this purpose, and it is explained in what respects they are to be preferred to other types. A physical equivalent circuit, i.e. one in which the elements can easily be expressed in terms of the physical quantities governing transistor action, is then worked out. It is used to investigate how the h parameters for the common-emitter configuration depend on the D.C. operating point (biassing) and on temperature. A more detailed explanation of how the Early effect is allowed for in the physical equivalent circuit is set out in an appendix.

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.

2447: H. Zijlstra: Maximum energy product of crystal-oriented polycrystalline Ticonal G (Alnico 5) magnet steel (J. appl. Phys. **27**, 1249-1250, 1956, No. 10).

Measurements and calculations relating to the effect of crystal orientation on the magnetic properties of Ticonal G.

2448: J. A. W. van der Does de Bye: Analysis of the scintiscanning problem (Nucleonics **14**, No. 11, 128-134, Nov. 1956).

"Scintiscanning" implies the mapping, by means of a scintillation counter, of parts of the human body which have a selective uptake of certain chemical compounds labelled with gamma-emitting tracers. This paper deals with the problems of collimator design and measuring procedure in order to arrive at the best possible picture with a given isotope quantity and a given measuring time.

2449: W. Verweij: Die Zündung explosiver Gasgemische beim Bruch von nicht brennenden Leuchtstofflampen (Z. angew. Phys. **8**, 521-530, 1956, No. 11). (The ignition of explosive gas mixtures by the breakage of non-burning fluorescent lamps; in German.)

The mechanism of gas-dynamic (shock-wave) ignition occurring with the breakage of gas-filled tubes has been investigated. Whether ignition of a given atmosphere occurs or not depends on the pressure and the molecular weight of the rare-gas filling of the tube. The pressure and temperature explosion limits were determined for various pressures and molecular weights of the rare-gas filling; the atmospheres investigated were air mixtures containing hydrogen (45%), methane (10%) and ordinary main gas (20%). The explosion limits observed are explained in terms of the calculated values of pressure and temperature occurring in the tube. The experiments were carried out at gas concentrations showing the lowest explosion limits; these are the most dangerous concentrations for gas-dynamic explosions but rarely occur in practice. It is shown that mixtures of air with main gas and with methane do not explode when a fluorescent tube breaks owing to the extinguishing effect of the cloud of fluorescent powder which is blown off the walls by the inrushing gas.

2450: S. Woldring: Continuous non-bloody recording of arterial pressure in man (Acta physiol. pharmacol. neerlandica **5**, 250-251, 1956, No. 2.)

Note on an apparatus employing the "floating arterial wall" for recording blood pressures in man.

2451: C. M. van der Burgt: Performance of ceramic ferrite resonators as transducers and filter elements (J. Acoust. Soc. Amer. **28**, 1020-1032, 1956, No. 6).

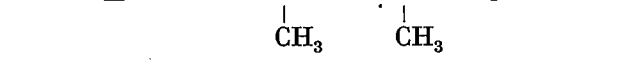
In recent experiments the compositions and methods of preparation of various ferrites were varied with the aim of achieving optimum mechanical and piezomagnetic performance. Essential improvements were obtained by small cobalt substitutions in the chemical composition and by suitable modifications of the mechanical and thermal treatments. In continuous wave operation, appropriate ferrites can excite radiation intensities up to at least 10 W/cm^2, with an electroacoustic efficiency of more than 75% up to 50 kc/s. They are therefore highly promising for underwater transmission and processing, in spite of their brittleness (ultimate tensile strength up to about 6 kg/mm^2). Wide-band hydrophones with linear frequency response and high pressure-sensitivity can also be built from such ferrites having magnetomechanical coupling coefficients at remanence up to almost 0.25, like Permendur. The total variation of the mechanical resonant and antiresonant frequencies of ferrite filter elements in the temperature range from +20 to +50 °C, being 0.1-0.25% for existing commercial ferrites, has been lowered to 0.03% and less. Since the mechanical quality factors are usually much better than 2000, such ferrites can be profitably applied to the construction of electrical and electromechanical band filters.

2452: J. L. J. van de Vliervoet, P. Westerhof, J. A. Keverling Buisman and E. Havinga: Studies on vitamin D and related compounds, VI. The synthesis and properties of dihydrotachysterol$_3$ (Rec. Trav. chim. Pays-Bas **75**, 1179-1186, 1956, No. 9-10).

Whereas dihydrovitamin D_3 I and dihydrovitamin D_3 II can be isolated from the mixtures obtained on reduction of cholecalciferol with sodium and alcohols, attempts to obtain dihydrotachysterol$_3$ in a crystalline state from such mixtures were unsuccessful. Pure dihydrotachysterol$_3$, however, could be obtained by reducing tachysterol$_3$ with an excess of lithium in liquid ammonia, concentrating by chromatography on alumina, esterification with phenylazobenzoyl chloride and saponification of the ester. Ultra-violet and infra-red absorption spectra are given. The capacity of dihydrotachysterol$_3$ to raise the calcium level of the blood (plasma) proved to be about twice as large as that of dihydrotachysterol$_2$.

2453: H. D. Moed and J. van Dijk: Synthesis of β-phenyl-ethylamine derivatives, IV. A new vasodilator (Rec. Trav. chim. Pays-Bas **75**, 1215-1220, 1956, No. 9-10).

The synthesis and properties of a new vasodilating agent are described:



2454: P. Westerhof and J. A. Keverling Buisman: Investigations on sterols, VIII. Some hitherto unknown irradiation products of ergosterol (Rec. Trav. chim. Pays-Bas **75**, 1243-1251, 1956, No. 9-10).

From the "toxisterol$_2$" mixture obtained by prolonged U.V. irradiation of ergosterol, two or possibly three hitherto unknown compounds have been isolated in the form of crystalline esters.

2455: H. Bremmer: Asymptotic developments and scattering theory in terms of a vector combining the electric and magnetic fields (I.R.E. Trans. **AP-4**, 264-265, 1956, No. 3).

The Maxwell equations for an inhomogeneous medium are expressed in terms of a complex vector combining the electric and magnetic fields. It is shown how this vector itself and the corresponding complex conjugated vector are coupled due to the inhomogeneity of the medium. The coupling terms consist of two contributions which may be attributed to scattering effects and to effects produced by internal or gradient reflections respectively. The above-mentioned vector combination proves to be particularly suitable for a formulation of the geometric-optical expansions applied in the theory of Luneberg and Kline.

2456: A. A. van Trier: Some topics in the microwave application of gyrotropic media (I.R.E. Trans. **AP-4**, 502-507, 1956, No. 3).

The Faraday effect of plane and guided waves is reviewed in the first sections of this paper. A cavity technique is then described for measuring Faraday rotations in a circular waveguide with a coaxial ferrite pencil. Some experimental results are discussed, including the evaluation of the permeability tensor components, the relation between Faraday rotation and pencil radius, and ferromagnetic resonance in circularly polarized waves. The problem of the rectangular gyrotropic waveguide is taken up in the last section of the paper. A simple method of successive approximations is described and applied to the case of the square waveguide.

2457: E. G. Dorgelo: Technologie von Magnetrons und Klystrons (Vakuum-Technik **5**, No. 8, 177-189, 1956; in German).

In the first part of this paper (see these Abstracts No. 2378) an introduction on the general construction of a few typical magnetrons and klystrons was given and the following technological problems were dealt with in more detail: the influence of dimensional tolerances on the behaviour of the valve; losses in the internal cavities and the coupling system; thermal and secondary electron emission from the cathode, suppression of unwanted emission and temperature stability under back-bombardment; factors influencing life of the valve. The second part of the paper now gives a survey of manufacturing techniques. Special attention has been given to different types of cathode. An L-cathode, specially designed for use in magnetrons, is described.

2458: J. A. Haringx: Enkele essentiële elementen bij ver doorgevoerde automatisering (De Ingenieur **68**, E.126-E.131, 1956, No. 50). (Essential elements in high-level automation; in Dutch.)

Three different levels of automation are to be distinguished. They are characterized by *a*) mere programming, *b*) programming and feed-back, *c*) decision, programming and feed-back. A superficial treatment of an electronically controlled machine tool shows how mechanical actions can accurately be automatized by the introduction of the electronic computer.

2459: J. Bloem and F. A. Kröger: The interstitial diffusion of Cu in PbS single crystals (Report of the Meeting on Semiconductors, Physical Society/British Thomson-Houston Ltd., Rugby, April 1956, pp. 77-81).

The diffusion of Cu into PbS crystals was studied at low temperatures (100-400 °C), where the self-diffusion of Pb or S may be neglected. The diffusion rate depends on the previous history of the crystals and on the ambient gas atmosphere during the diffusion. The activation energy for interstitial Cu diffusion is found to be 0.3 eV.

2460: H. Bremmer: Numerical analysis of magnetic lens parameters on a theoretical basis (Proc. 3rd Int. Conf. Electron Micros. London, 1954, pp. 89-92).

An account is given of the present state of a systematic survey of the properties of magnetic electron lenses. Paraxial parameters as well as aberration coefficients are derived from numerical solutions of the paraxial-ray equation.

2461: A. C. van Dorsten: Electron irradiation of specimens (Proc. 3rd Int. Conf. Electron Micros. London, 1954, pp. 172-173).

Specimens too thick to be studied in the electron microscope without sectioning have been exposed to an external beam of 700 kV electrons. Increased transparency with preservation of structure has been observed in some cases.

2462*: G. W. Rathenau and G. Baas: Emission electron microscopy (Proc. 3rd Int. Conf. Electron Micros. London, 1954, pp. 387-390).

A description is given of the method and apparatus used in metallographic studies of metal surfaces by emission electron microscopy. See also Philips tech. Rev. **16**, 337-339, 1954/55 and **18**, 1-10, 1956/57, No. 1.

2463*: H. B. Haanstra: The use of surface-active materials in electron microscopy, particularly for surface replicas (Proc. 3rd Int. Conf. Electron Micros. London, 1954, pp. 489-490).

Surface-active materials have been used in preparing replicas, in particular for positive replicas, of wet surfaces. The method is an improvement of the methyl methacrylate-silica method of Brown and Jones.

VOLUME 19, 1957/58, No. 4 pp. 109-144 Published 16th October 1957

Philips Technical Review

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

THE FUNCTION OF ADDITIVES IN TUNGSTEN FOR FILAMENTS

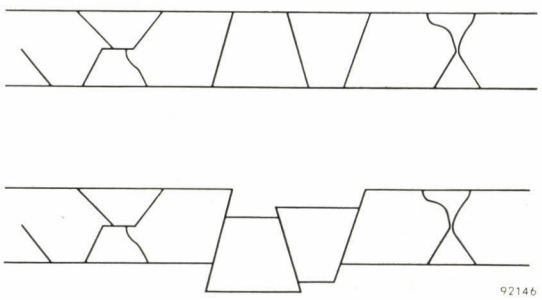
by J. L. MEIJERING and G. D. RIECK.

621.3.032.321:669.27

The production of coiled tungsten filaments which do not deform when incandescent is one of the most important features of modern incandescent-lamp manufacture. It was discovered some decades ago that such coils kept their shape excellently when small amounts of certain oxides were added to the tungsten at the right stage in the manufacturing process. However, even today the details of the role that these additives (dope) actually do play in the filament are still unknown. This article gives some considerations on the subject and describes a number of experiments that cast some light on the matter.

Ever since it first became possible to prepare metallic tungsten in a ductile state and to draw wire from it, the incandescent lamp industry has come to be based almost entirely on the use of this metal as its filament material. The development of processes that made this possible dates from 1909. Since then there has been only one revolutionary step in the evolution of the incandescent lamp, namely the adoption of gas-filled bulbs. At the same time coiled and coiled-coil filaments were introduced. This improvement, which was based on Langmuir's theoretical considerations (1912) and led to an appreciable improvement in the luminous efficiency [1]), at first looked as though it would come to nothing because of the behaviour of the drawn tungsten filaments.

When a coiled filament is heated to high temperature, it readily deforms under the force of gravity ("sagging"). A phenomenon which sometimes accompanies this, and is liable to detract from the shape and life of an incandescent tungsten coil, is the shearing [2]) along the boundaries of those tungsten crystals which bridge the whole diameter of the wire with relatively short boundaries ("offsetting", *fig. 1*). This article describes how these problems can be dealt with, and presents a picture (partly based on experiments conducted by the authors) of what happens during and after the drawing of the wire. We shall first, however, briefly describe how a coiled filament is manufactured.



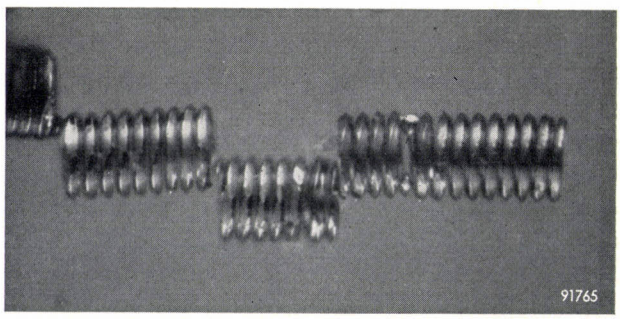


Fig. 1. *Above:* Schematic diagram showing "offsetting" in a tungsten filament along crystal boundary planes which are roughly at right-angles to the wire axis.
Below: Photograph of a coiled tungsten filament in which offsetting has occurred at various points.

[1]) Cf. W. Geiss, Philips tech. Rev. **6**, 334-342, 1941.
[2]) It will be clear to the reader that the word shearing is used here in a different sense than in the study of plastic deformation and that there is no question here therefore of the sliding of a crystal along one of its glide planes.

Tungsten wire is fabricated from bars by swaging and drawing. Tungsten has a very high meltingpoint (a necessary characteristic for incandescent filaments); the bars are therefore made not by melting but by the method of powder metallurgy. A very fine metallic formed is prepared by reduction of the oxide with hydrogen. This powder is pressed into bars under an extremely high pressure (10-40 kg/mm^2). The bars are given sufficient coherence for handling by pre-sintering in a furnace. By passing a current through the bars, a temperature just short of the melting point can be reached, this taking the sintering process a good deal further (as can be seen from the increase in density and other characteristics). The bars are then strong enough to be put through a kind of forging process: the white-hot bars are swaged, i.e. machine-hammered into thinner, much longer lengths, this process also further increasing their density. The bars are now drawn into wire through hard-metal dies; wire of the smallest diameters is obtained by drawing through diamond dies [3]). When at last the wire is about 100 μ in diameter, the density is 19.2 g/cm^3 and has thus risen almost to that of the bulk metal (19.3). This wire can now be wound round a wire mandrel to form a coil. If this coil is wound, still on its mandrel core, around another mandrel, a coiled coil is obtained. By heating to incandescence, the tungsten coil becomes "set" in its new form and does not spring back when the mandrel cores are dissolved away chemically.

Owing to the strong deformation to which the metal crystals are subjected during drawing of the wire, these crystals become stretched out along the axis of the wire. These long drawn-out crystals are known as "fibres" and the fibre structure that forms is shown in *fig. 2*. When such a wire is heated to high temperature in the incandescent lamp, it recrystallizes, i.e. new crystals form in the aggregate of deformed crystals. Some control over this recrystallization is necessary if the coil is to be prevented from sagging and offsetting. When *pure* tungsten is used, fairly large numbers of new crystals form on recrystallization; such a filament soon begins to sag. Large crystals which take up almost the whole diameter of the filament, and have boundaries substantially perpendicular to the filament axis, may cause offsetting.

It might have taken years of intensive research to discover a process that led to good filaments, had not chance intervened: it was noticed that

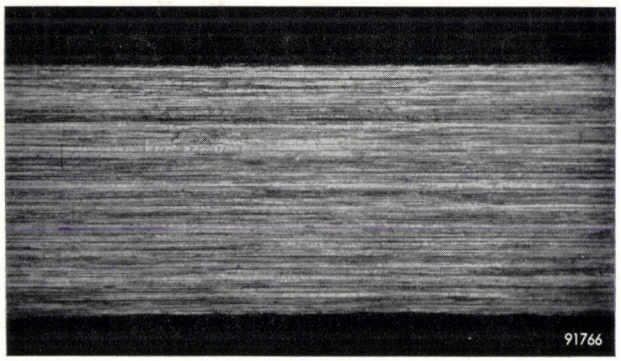


Fig. 2. The "fibre structure" of a drawn tungsten wire which has not yet recrystallized. Magnification 265×.

tungsten exhibited particularly good non-sag properties in the recrystallized filament or coil, if the metal had been prepared from raw material (WO_3) which had been heated in a Battersea crucible [4]). It turned out that an extremely small amount of the material of the crucible found its way into the tungsten oxide, and that the active components of the contaminant were the oxides K_2O, Al_2O_3 and SiO_2. Addition of these substances to the tungsten at the appropriate stage in its manufacture, was found to result, during recrystallization, in the formation of longer boundaries between the various crystals; moreover, these boundaries made, in general, smaller angles with the wire axis, often assuming sharp arrow-head shapes (*fig. 3*). This remarkable effect is far from being the only influence that the additives ("dope") have on the properties of tungsten wires. A discussion of these effects and of the innumerable experiments that have been carried out in this field would fill a book, however. We shall confine ourselves therefore to examining the long, arrow-shaped boundaries which evidently prevent sagging and offsetting; we shall also put forward an explanation of the effect of the additives on the formation of the crystal boundaries.

Before going any further it should be remarked that the additives, which are added to the WO_3 in small amounts (a few per cent by weight), largely disappear during the subsequent manufacturing process. Only about 0.01% of Si and Al can still be detected in the filament as it goes into the lamp. Nevertheless these very small amounts appear to be adequate to influence recrystallization in the desired manner.

[3]) Cf. J. D. Fast, Philips tech. Rev. **4**, 309-316, 1939.

[4]) This was a type of crucible of Al_2O_3-rich earthenware made by the firm "Battersea" and formerly much used in chemical plants and laboratories.

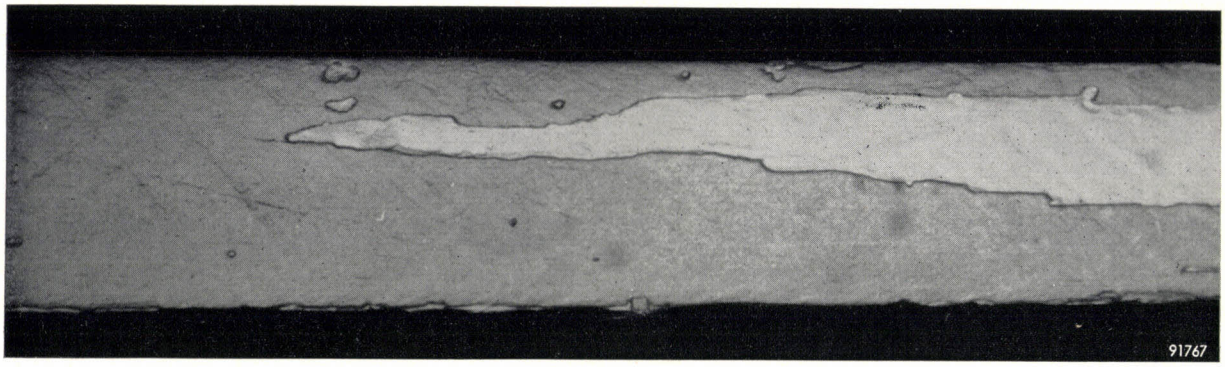


Fig. 3. Arrow-shaped boundary between two crystals in a recrystallized tungsten wire. Magnification 200×.

The form in which the additives are present and their influence upon the recrystallization process

According to present ideas, the substances added are thought to form compounds (e.g. potassium silicotungstates, or perhaps tungsten bronzes) which are fairly stable, possible survive the reduction process and do not completely evaporate at the high sintering temperature. These compounds, or what is left of them in the form of decomposition products, are present as impurities that are not soluble in the tungsten, but form a separate phase. It is plausible to assume that these slag-like impurities are spread out — like the metal itself — along the wire during the drawing process, and that they occur in the wire along lines or planes or as tubules. We shall now describe a number of model tests which illustrate how such fibrous and tube-shaped structures can make for the formation of the desirable long crystal boundaries [5]).

Model tests

Let us imagine a length of wire containing impurities in tubular form, i.e. we picture the wire as a bundle of thin-walled tubules, the wall of each tube being made up of the residues of the additives and the space between them consisting of fine-grained tungsten. Owing to the irregular manner in which the tubules arise, they will be blocked here and there by cross-walls (dams), while at other places there will be holes (leaks) in the walls. In the actual wire the situation will be rather more involved, but will in principle be the same.

Let us now assume that somewhere in the "tungsten" part, the nucleus of a new crystal forms and begins to grow. In the tubule in which it is contained, the crystal nucleus grows unhindered to form a small crystal, but at the walls and dams in the tubule its growth is impeded. However, the crystal can grow out through a leak in the wall into the adjacent tubule and can expand throughout the whole of that tubule up to the dams; it may then grow further through leaks in this tubule.

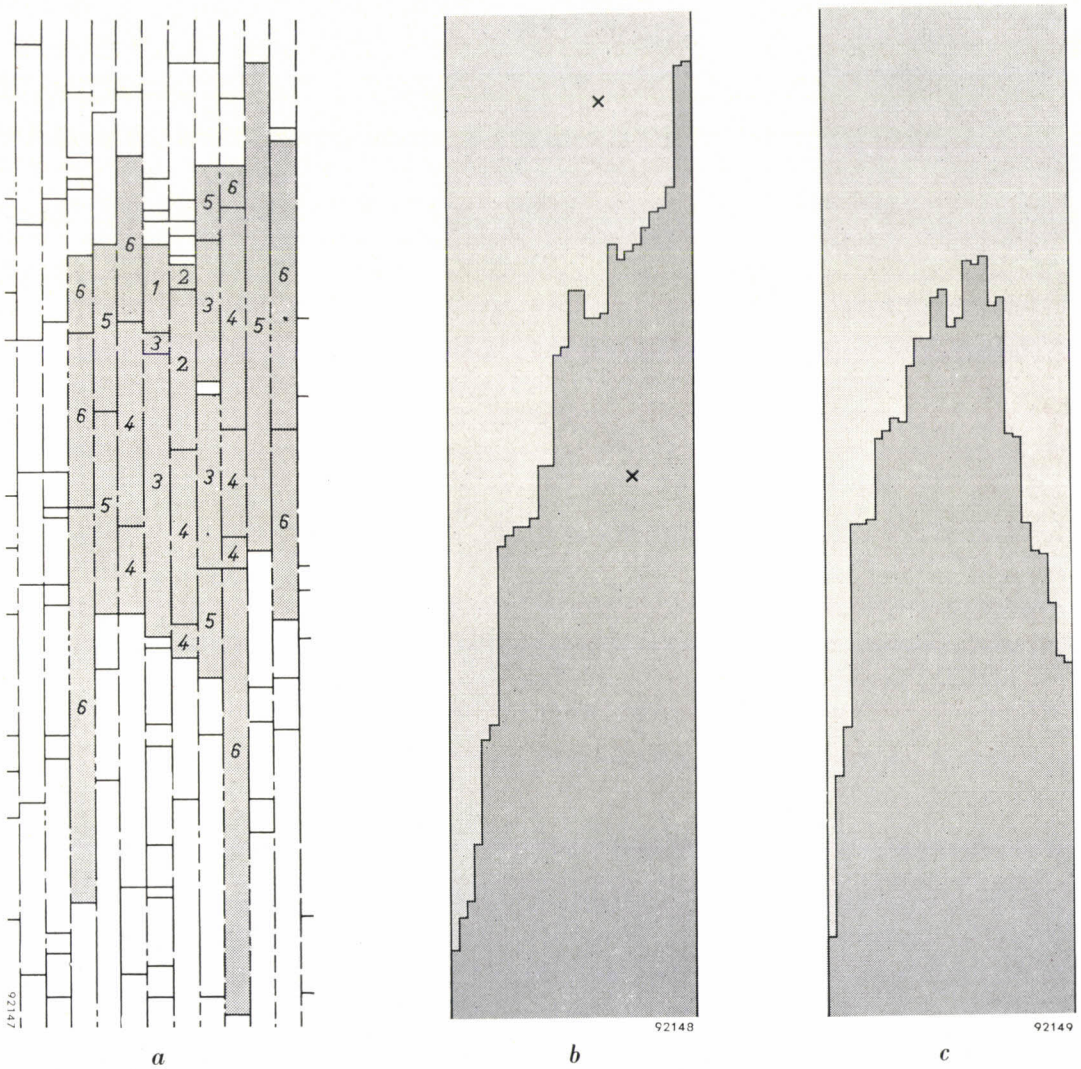
Minor holes in the walls may nevertheless not act as leaks: owing to crystal-boundary energies, it is only fairly large leaks that will be passable (cf. a liquid with a high surface tension that does not leak from a porous vessel). Also, a row of pores can have an impeding effect just as a row of foreign particles.

If we restrict our representation to two dimensions, we can imitate what happens by constructing a simple model as follows. A large number of parallel lines, representing the tubule walls in the wire, is drawn on a sheet of paper. At random, dams are drawn in indicating where the tubes run to a dead end, and leaks are inserted by interrupting the long lines at various places. In the first series of experiments, we make the number of dams and leaks equal; in subsequent series of experiments we can take unequal numbers of dams and leaks.

With such a diagram we can perform the following "recrystallization" experiments. We create a crystal nucleus at each of two places and allow them to grow along the tubules and through the leaks according to certain rules. Each nucleus is permitted to advance a certain amount in turn: just how far is laid down by "rules", which may be different for each group of experiments. Thus, for example, we can allow each crystal in turn to fill up completely the section of tubule in which it started; it then has to wait until its next turn before it can occupy a neighbouring tubule via a leak. This corresponds to the assumption that the passage through a leak takes a relatively long time. Another possibility is that the crystals are allowed to grow at a constant rate. Still other possibilities, of course, exist. *Fig. 4a* shows a "crystal" in course of growth according to the first-mentioned of these rules.

[5]) J. L. Meijering, Modellversuche über das Entstehen des Stapeldrahtgefüges in Wolfram, published in Warmfeste und korrosionsbeständige Sinterwerkstoffe, 2. Plansee-Seminar 1955, Reutte/Tirol, pp. 305-312 (published 1956).

Two crystals growing in such a manner will finally meet at one or other boundary line. It is interesting that these boundary lines show a strong similarity to the crystal boundaries that occur during the recrystallization of tungsten filaments prepared from "doped" tungsten. The long boundaries making a small angle with the longitudinal axis of the wire, and the arrow-shaped boundary lines can be recognized. Figs. 4*b* and 4*c* are examples of the results obtained in this "recrystallization on paper". By varying the number of leaks, dams and crystal nuclei present per unit area, structures can be obtained which resemble the recrystallization patterns of both the doped wires (long crystal boundaries) and the pure metal wires (roughly equidimensional crystals). According to these paper-model experiments not only the boundary between two new crystals, but also that between a crystal and



a *b* *c*

Fig. 4. Paper-model experiments on the recrystallization of tungsten wire. The system of strands and tubules running lengthwise along the wire and made up of foreign particles assumed to be present in the wire, is represented in two dimensions by a system of vertical parallel lines, two typewriter spaces apart. Here and there these lines are broken ("leaks"), in other places lines across the tubes indicate blockages ("dams"). These were marked at random in our experiments with the aid of a typewriter and a table of random numbers, such as can be found for example by throwing a die or by any other system in which chance is decisive. We proceed as follows: reading through the list of random numbers, we type a mark on the page of vertical lines whenever a certain number, say 2, occurs. For all other numbers we strike the spacer. If we are *between* two lines, the mark made is a dash (—) and in this way we get a "dam". If we are *on* a line, we strike another character (e.g. =) which represents a leak at that point. (This gives approximately equal numbers of leaks and dams; it is easy to modify the "rules" so as to give other ratios of leaks to dams.) In the experiments crystals were permitted to nucleate at two random points in the diagram, each crystal being allowed "to make a move" in turn.

a) The growth of one such crystal. Each turn the crystal is allowed to occupy a complete section of tubule. The numbers indicate which parts of the tubules were filled in a particular turn. The shaded area indicates the shape of the crystal after six turns. In this model there were twice as many leaks as dams.

b) Oblique boundary between two crystals, as frequently found in these experiments. The two crystals grew from the points marked ×.

c) Arrow-shaped boundary. The crystal nuclei were situated at points outside the figure.

the fibre structure that has not yet recrystallized are of the "long boundary" type. This is in accord with what is actually observed when a crystal runs to a dead end in the fibre structure. Such a situation can arise for example when the temperature is just below the recrystallization temperature, rising above it only in a part of the wire. This is easily achieved by slightly etching away part of the wire, so that the current density is not everywhere the same (*fig. 5*). When a specimen of such a wire is viewed under powerful magnification, a stepwise pattern can be seen where the crystal growth into the "fibres" has halted.

This demonstrates how impurities can inhibit the growth of the crystals — in particular, growth at right-angles to the wire axis mainly owing to the fact that the impurities are stretched out along the wire — and in this way promote the formation of long crystal boundaries. It will be obvious that it is chiefly the least volatile of the additives, such as SiO_2 and Al_2O_3 or their compounds, that exercise this function.

The structure of the recrystallized tungsten wire

We have contended that it is highly probable that the residues of the additives are present in the drawn wire as elongated strands which exert a strong influence on the recrystallization of this wire. This view is confirmed by a second series of experiments, now to be described, which show that the additives are still present in the form of tubules or strands when the recrystallization process is over [6]).

The structure of the wire was investigated by making Laue diffraction photographs, using a very narrow beam. The latter was obtained by means of a 2 cm long lead-glass collimating capillary of about 50 μ bore. With this arrangement very small differences in orientation between crystallites can be detected. The very small diameter of the X-ray beam makes for a high resolving power (about one minute of arc); moreover the irradiated spot is so small that it covers only a few crystallites. *Fig. 6a* shows a Laue photograph (taken with a normal beam, not specially narrow) of a crystal of rock-salt orientated arbitrarily with respect to the X-ray beam. Fig. *6b* shows how such a photograph appears when the beam falls on two crystals having a small angle with respect to each other (doubling of Laue spots).

[6]) G. D. Rieck, Fragmentation in tungsten crystals, Acta metallurgica **4**, 47-51, 1956 (No. 1).

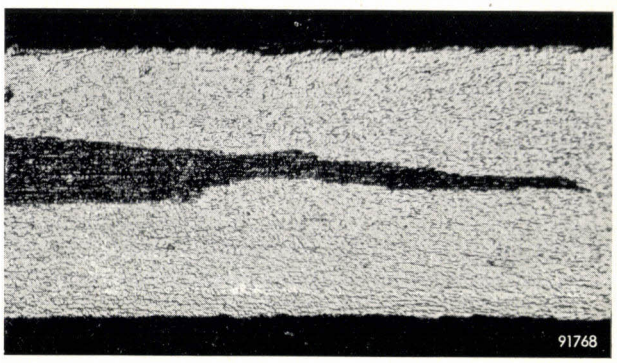
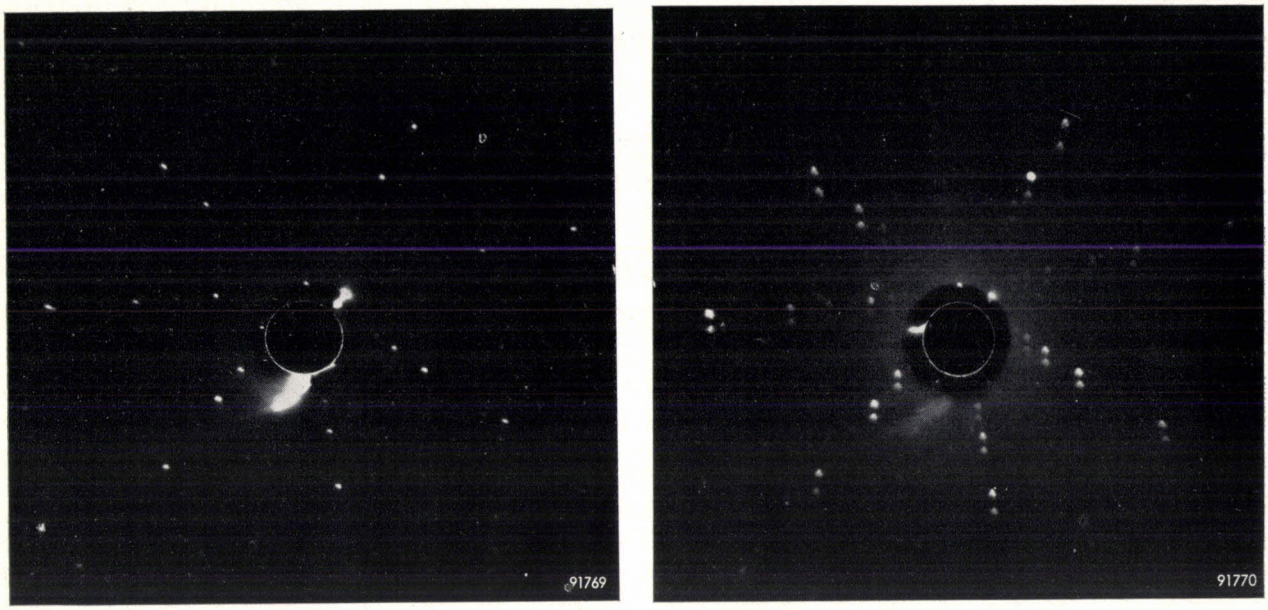


Fig. 5. Photograph of the boundary between a new crystal and an unrecrystallized region of the wire (the fibre structure). The required temperature gradient was obtained by slightly etching away part of the wire. The darker region in the photograph is the original fibre structure, the lighter region is recrystallized metal. Magnification 160×.

If back-reflection photographs are made of recrystallized tungsten wires at places where the whole cross-section is taken up by only one crystal, it is often found that one spot in the diagram obtained is split into two or more smaller spots. More than half the crystals examined were found to exhibit this splitting of the spots, and it was noteworthy that this splitting usually corresponded to rotation about an axis *parallel* to the wire axis. *Fig. 7* shows a few examples of parts of such diagrams [7]). From the distance between the spots, the angle between the irradiated crystallites can be calculated. Angles of about 2′-30′ are commonly found. From the surface area irradiated and the number of spots found, moreover, it is possible to make an estimate of the size of the crystallites involved. The diameter found varies from 20 to 70 μ.

It may be remarked at this stage that the axial rotation of the crystallites with respect to each other fits in well with the hypothesis of strands or walls of impurities running parallel to the wire axis, and with the picture of recrystallization that has just been discussed. For it is reasonable to expect that the parts of a crystal on either side of a wall will often show small differences in orientation. The two parts of the crystal will be in contact only at one point, namely at the leak from which the second part began its growth. The geometry of the wire then lends greatest probability to a difference in orientation corresponding to rotation about the wire axis, in accord with the rotational symmetry of the fibre structure of the wire.

[7]) Splitting of the spots also occurs as a result of trivial effects, viz. 1) if the X-ray beam strikes the film at an angle and the film has emulsion on both sides; on developing the film two spots appear, on the front and back of the film, which do not coincide; and 2) if the lead-glass capillary is not sufficiently narrow, so that the local differences in brightness of the anode are manifested in each spot.



a *b*

Fig. 6. A Laue photograph of NaCl. When white X-rays (continuous spectrum) fall upon a crystal, the various planes in the crystal occupied by atoms or ions diffract this radiation. The pattern which is formed with a given angle of incidence of the X-ray beam, is determined by the position of the various lattice planes in the crystal. Laue diagrams can be made either with small angles between the incident and reflected beams, the film being mounted behind the crystal (transmission photograph), or with large angles, the film being mounted in front of the crystal (back-reflection photograph). In the latter case the film has an aperture to let through the incident X-ray beam.
a) Back-reflection Laue pattern of a single crystal. *b*) Back-reflection Laue diagram of two crystals which are turned through a small angle with respect to each other. (These photographs were not made with a very fine beam.)

Bending test

The wires containing the crystallites just discussed were fixed horizontally on a specimen holder, at right-angles to the X-ray beam. The following experiment was then performed on these wires. They were bent through an angle of 90°, in a plane at right-angles to the X-ray beam, with a radius of curvature of (for example) 2 mm (the diameter of the wire was 180 μ). The wires were then straightened as far as possible. The experimental set-up made it possible to ensure that the X-ray beam impinged on exactly the same spot in the previously examined part of the wire, which was also the middle of the bent region.

As is well known, a deformed crystal gives a Laue diagram with elongated spots (asterism) [8]). The direction in which the spots are elongated is related to the direction in which the lattice plane in question is curved. When the tungsten crystals were bent back, this asterism was found largely to disappear as far as the spot elongation parallel to the axis of the wire was concerned; the spots were usually little changed in a horizontal direction after the experiments. In the vertical direction, however, something had happened. Spots that had displayed some degree of splitting earlier, were more decidedly split, and in spots where no splitting had previously been observed, this was frequently found after the bending test. The angular differences were again of the order of a few minutes, but the number of spots was often bigger, corresponding to particles of about 7 μ diameter. The phenomenon that the effect of bending was more persistent in the "vertical" than in the "horizontal" direction can now be explained on the assumption that strands of impurities are present in the crystal investigated. The explanation is in terms of lattice defects (dislocations) that arise as a result of deformation [9]).

When a substance is bent plastically, edge dislocations are formed. An edge dislocation can be imagined as being caused by the presence of an extra atomic half-plane, in this case inserted into the crystal lattice from the outside of the bend

[8]) See, for example, W. G. Burgers, Philips tech. Rev. **5**, 157-166, 1940.

[9]) See, for example, Philips tech. Rev. **15**, 246 and 286, 1953/54.

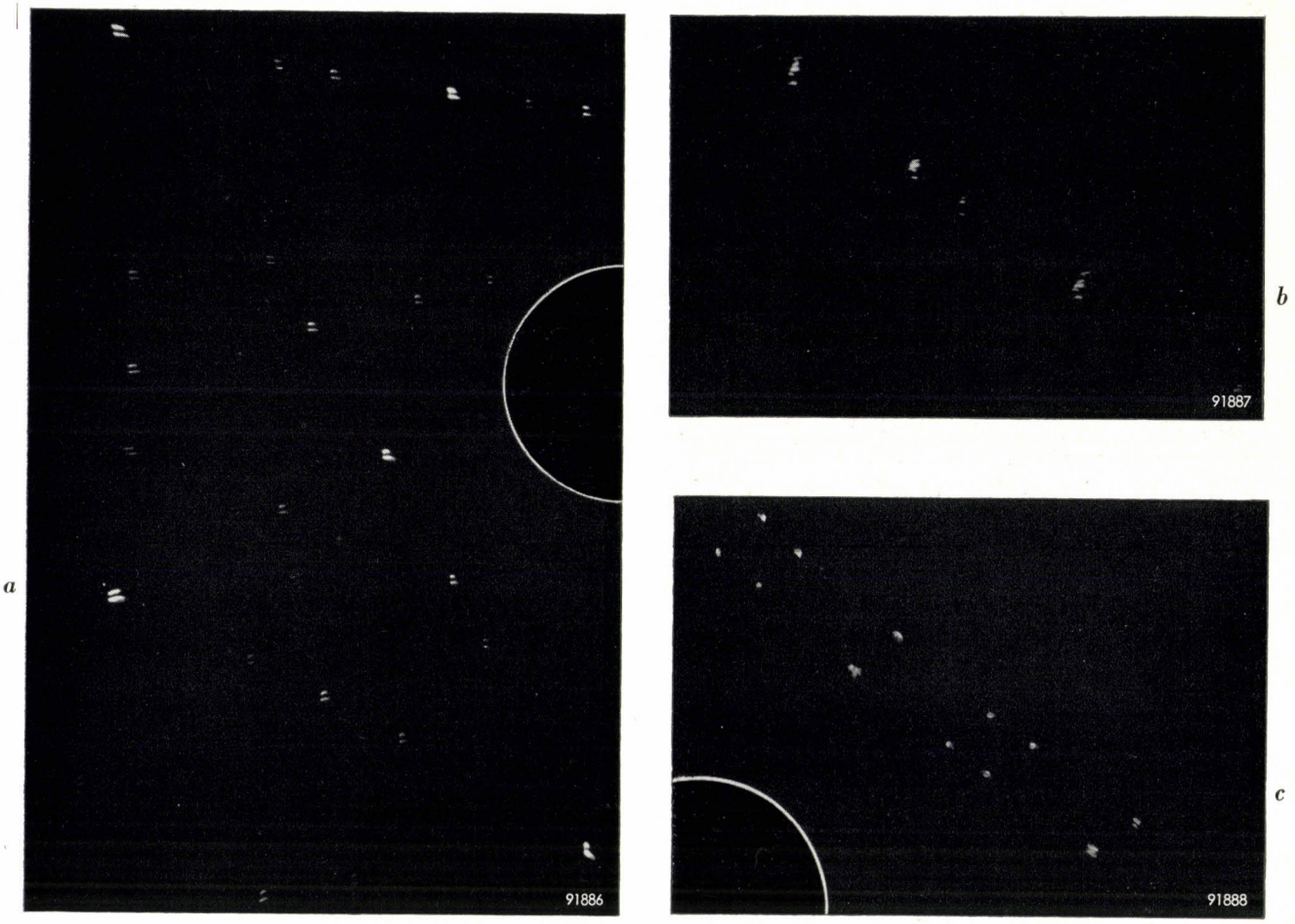


Fig. 7. Laue photographs of recrystallized tungsten wires. A splitting of the spots is observed, this being due to small differences in the orientation of the crystallites: *a*) two-fold splitting (three times actual size); *b*) five-fold splitting (four and a half times actual size); *c*) the spots in this photograph originate from two crystals which are at an angle of 2° and are each separately re-split (three times actual size).

(*fig. 8*). If the wire is now bent in the opposite direction, dislocations are produced which can be thought of as resulting from extra atomic half-planes inserted from the opposite side of the crystal. The extra half-planes of these two kinds may often compensate each other, provided the mobility of the dislocations is sufficiently great. The mobility of a dislocation is markedly reduced if the edge of the extra atomic half-plane, i.e. the dislocation axis, comes up against some foreign atoms (impurities) in the crystal lattice. If the dislocation axis coincides with a *row* of foreign atoms, then the dislocation may be rendered immobile and no longer able to combine with an opposing dislocation (which in its turn may also be locked against another, parallel row of foreign atoms). However, dislocations whose axes are at right-angles to such rows of impurities will be practically unimpeded.

We have seen that in the tungsten crystal the

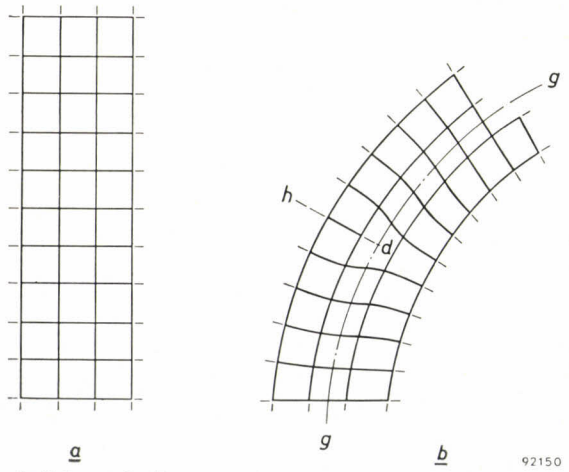


Fig. 8. Schematic diagram of part of a crystal lattice in which shearing has occurred as a result of bending, so that an edge dislocation has formed. *a*) The lattice prior to bending. *b*) The same section of the lattice after bending. Shearing has occurred along the glide plane *g*; this is equivalent to the insertion of an extra half-plane of atoms at *h*. The edge of this plane, i.e. the line at right-angles to the plane of the drawing through the point *d*, is the dislocation axis.

changes in the spots resulting from deformation purely in the plane of the bent wire (as in fig. 8), are restored to normal when the wire is bent back. The corresponding dislocation lines are then at right-angles to the axis of the wire and there is evidently nothing to impede their mobility. However, the (secondary) deformation corresponding to the change in the shape of the cross-section on bending the wire, which manifests itself in a splitting of the spots in a direction at right-angles to the wire (see fig. 7), does not disappear. The dislocations corresponding to this deformation are mainly parallel to the wire axis and evidently they are partly blocked. As we have seen, such a locking of these dislocations is to be expected where there are rows of foreign atoms likewise running parallel to the axis of the wire. This is thus in very good agreement with the view that there are still traces of impurities present in the recrystallized tungsten crystals in the wire, these impurities being arranged in lines parallel to the wire axis.

The rows of foreign particles need not, of course, be one atom thick; to explain the difference in mobility of both types of dislocations, it is sufficient to assume that these particles are distributed more extensively along the axis of the wire than at right-angles to it.

We thus find confirmation for the picture of the impurity distribution obtained from the experiments with paper models, which gave an acceptable explanation of the way in which long and arrow-shaped boundaries can arise between the crystals in a tungsten filament.

Other conceivable explanations of the occurrence of long boundaries

When two crystals grow, for example in a piece of metal, at a certain moment they will meet. It is then possible that one of the two crystals will continue to grow at the expense of the other [10]); alternatively they may maintain one common boundary. In general it can be said that when the boundary-surface energy is low, there is little tendency for the boundary to move by the growth of one crystal at the expense of the other. In such a case the boundary can in general be a long one (since its total energy will then still be low). Such a long boundary could for example occur between twinned crystals [11]).

[10]) In the case of iron alloys this phenomenon has actually been followed, as it occurred, under an emission electron microscope, cf. Philips tech. Rev. **16**, 337-339, 1954/55, and **18**, 1-10, 1956/57 (No. 1).

[11]) A twinned crystal consists of two parts which are in a certain symmetry relationship to each other: they are either mirror images of each other with regard to a common lattice plane, or they are turned through 180° with respect to each other about a two-fold axis of symmetry (both features may be present at once). The mirror plane usually has simple indices and is *not* one of the planes of symmetry of the crystal structure. In general it does not coincide with the interface, which need not be flat.

Also, crystals which have almost the same orientation, can easily exist side by side and may possess long crystal boundaries.

It might be that the crystals having long boundaries in tungsten wires are in one or other of the two relationships described above. It is of course *a priori* improbable that two crystals which nucleate some distance apart and grow towards each other, would possess a twin-crystal relationship, but it can well be imagined that such a relationship might occur if one of the two crystals were to arise from a nucleus that was formed on the growth front of the other. Disregarding special orientations it is further not unreasonable to suppose that such a crystal on the growth front of another, would locally affect the growth of the latter and by so doing contribute to the formation of a curiously shaped boundary; such a situation has actually been found in aluminium [12]).

X-ray diffraction patterns showed, however, that the angles between tungsten crystals with long boundaries can have various values; also, tungsten crystals bordering on each other were not usually found to have a twin relationship, though twinned crystals might under some circumstances occasionally occur.

We have investigated in the following manner whether two crystals which grow towards each other and are thus unable to influence each other, are able to form a "long" boundary.

"Shunt" experiments

It is possible to make a crystal boundary such that it is certain that the two bordering crystals are nucleated in the fibre structure at least $1^1/_2$ cm apart. This has been done as follows. The wire (in these experiments of diameter about 0.3 mm) is fitted with a sleeve ("shunt") of length, say, 15 mm (see *fig. 9*) between the points *A* and *B*. The "shunt" consists of six pieces of the same wire as that under test, cut to this length, and positioned around the wire; they are held in place by means

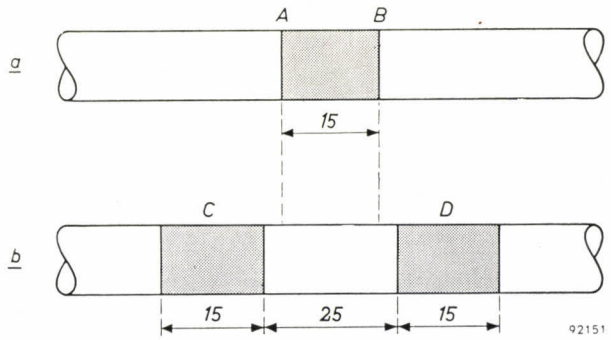


Fig. 9. Schematic diagram of the recrystallization experiments designed to obtain crystal boundaries with the certainty that the crystals involved had been nucleated far apart from each other. A sleeve ("shunt") was fixed around the wire. Within the shunt the wire is less strongly heated, so that recrystallization cannot occur there. The dimensions are given in mm. The "cold" sections are shaded in.
a) In the first heating the shunt envelopes the section *AB*. Crystals originating from the left or right cannot grow into the region *AB*.
b) In the second heating two shunts are mounted. The crystals whose growth was halted at *A* or *B* can now continue their growth. If crystals are found whose one extremity lies in *AB* and whose other extremity is beyond *C* or *D*, it is certain that these crystals have nucleated *outside* the section *AB* in the first heating. As a rule only one crystal boundary is found in the section *AB*, this being of the same type as those found after recrystallization under normal circumstances.

[12]) W. G. Burgers, Physica **9**, 987-995, 1942.

of a piece of thin wire wound round them. The shunted wire is mounted stretched and vertical in a tube through which hydrogen is passed. On heating the central wire to about 2250 °C by passing a current through it, the shunted section does not rise above 1000 °C and therefore does not recrystallize. After fifteen minutes, the wire is cooled and is then heated with *two* shunts (*C* and *D* in fig. 9). Finally, a longitudinal section is prepared for microscopic examination.

As a rule one crystal boundary is found in the section *AB* which represents the boundary between the crystals that were halted at *A* and *B* by the "cold" section *AB* during the first heating. If the one crystal extends to the left of *C* and the other to the right of *D*, we can be *certain* that they were formed left of *A* and right of *B* during the first heating, since the two "protective" shunts used in the second heating prevent any crystal that might nucleate between *A* and *B* (e.g. on the growth front of the "left-hand" crystal) from growing out to the right beyond *D*, swallowing up another crystal.

Complications can arise; for example a third crystal may form in *AB* or may already be present next to one of the two other crystals at *A* or *B*. But on various occasions there was only one boundary between *C* and *D*, and this was always of the same type as is found when a filament is heated to incandescence in the normal way. Nor were any twinned crystals found. In a few instances the second heating was extended to 10 hours at 2400 °C. This did not lead to different results, so that the boundaries evidently possess an appreciable stability.

From all this it must therefore be concluded that the simple orientation relationship postulated above does not exist. However, this does not mean that some orientation relationship does not exist. Even after recrystallization there is a preferred crystallographic axis parallel to the wire direction. But the crystals can be turned through any possible angle with respect to each other, about this axis. [13])

[13]) This will be discussed at greater length in an article to be published elsewhere.

The experiments also demonstrate that the long or arrow-shaped boundaries found cannot be explained by the growth of a crystal nucleus on the boundary between the recrystallized region and the fibre structure. On the contrary, the actual growth front (between a crystal and the fibre structure) is found to be irregular and capable of assuming the arrow shape (fig. 5).

Summary. Tungsten wire for the filaments of incandescent lamps is made from sintered tungsten bars by swaging and drawing. In this process the crystals are deformed and a fibre structure is produced. If a wire is heated to incandescence, recrystallization occurs and crystals are formed that are larger than those previously present. If such a wire is now to be suitable for use as a coiled filament, it must not sag or show offsetting. Some decades ago it was found that these effects can be avoided by adding certain substances (dope) at an early stage in the production of the wire. Long crystal boundaries are then formed during recrystallization.

With the aid of two-dimensional diagram experiments a plausible explanation is given of how the formation of long crystal boundaries is promoted by the presence of foreign particles arranged in the wire in the form of strands or tubules. A system of tubules were drawn on paper and leaks and stoppages introduced in them in a purely random manner. Two "crystals" are made to grow and fill the system of tubules, starting from two points in that system, and following certain rules. X-ray diffraction patterns of crystals in recrystallized wires confirm the picture obtained. If a crystal is bent and then straightened, the dislocations of opposite sign can compensate each other when their axes are at right-angles to the axis of the wire, but not when they are parallel to the axis of the wire. The foreign particles forming the tubules are identified with the residues of the dope stretched out along the wire during drawing.

Experiments in which unrecrystallized tungsten wires are locally heated to high temperature show that the long or arrow-shaped boundaries between tungsten crystals cannot be explained in terms of twinned crystals, nor in terms of re-nucleation (with random orientation) on the recrystallization front.

A TRANSMITTING TRIODE FOR FREQUENCIES UP TO 900 Mc/s

by P. J. PAPENHUIJZEN.

621.385.3.029.6:621.3.032.213.1

*New designs have made it possible to use triodes as amplifiers or oscillators at unprecedently high frequencies. This was recently illustrated by an article in this Review on the EC 57 disc-seal triode *), which delivers a power of several watts at 4000 Mc/s. Considerable progress has also been made with triodes for appreciably higher power ratings. The transmitting tube TBL 2/300 discussed in the present article — also a disc-seal triode, but with a cylindrical electrode system as opposed to the planar electrodes of the EC 57 — has an output of about 400 W at 470 Mc/s and about 150 W at 900 Mc/s. It meets a need that has become increasingly pressing of late.*

In recent years the frequency range from about 450 to 900 Mc/s (wavelengths from some 7 to 3 dm) has come more and more into prominence. To give only a few examples, decimetric waves have acquired great significance in aviation for the purposes of communication and navigation; mobilophone communications at frequencies between 460 and 470 Mc/s are growing rapidly in number; in television more and more interest is being shown in bands 4 and 5 (470-585 and 610-960 Mc/s); high-frequency heating with dm waves is steadily gaining ground in industry; and in Germany a frequency range in the neighbourhood of 461 Mc/s has recently been reserved for diathermy, in connection with investigations being carried out into the therapeutic value of dm waves. In addition, numerous military applications might be mentioned.

Triodes of conventional construction are, for various reasons, unsuitable for decimetric waves. If we attempted to use such tubes at such wavelengths we should find in the first place that the electron transit time — the time taken by the electrons to travel from cathode to anode — would be too long in relation to the period of the oscillation to be generated or amplified. In the second place, the wavelength would no longer be large compared with the length of the electrodes, so that the voltages at different points of the same electrode would show phase differences; this effect becomes noticeable as soon as the wavelength drops to about 10 times the effective electrode-length. In the third place, at frequencies of several hundred Mc/s the dielectric losses in the insulating parts of the tube would be considerable, and finally, the self-inductance of the lead-in wires would exercise an adverse influence [1]). These four effects (transit-time effect, phase shift along the electrodes, dielectric losses and stray self-inductance) are, as regards transmitting triodes, the main reasons for the loss of efficiency at higher frequencies; at a certain limiting frequency the efficiency finally becomes too low for the tube to be of any practical use.

In the course of the years several tubes of an entirely different type have been developed, tubes whose operation is based upon the finite transit time of the electrons. The most important of these "velocity-modulated" tubes are magnetrons, klystrons and travelling-wave tubes. These are the tubes that have been mainly responsible for opening up the decimetric, centimetric and even the millimetric wave-ranges for radio engineering. Most of these tubes contain one or more cavity resonators, so that in principle each type is suitable for use at only one frequency. There are, it is true, various methods of varying this frequency to a slight extent [2]), but only over a small range (a few %). By the use of external cavity resonators and by very accurately adjusting the supply voltage, it is possible to extend the frequency range of klystrons and travelling-wave tubes, but none of these types can be employed as a universal tube for covering the entire frequency range up to, say, 900 Mc/s.

By contrast, the more conventional tubes (triodes, tetrodes) all operate with external oscillator circuits (e.g. cavity resonators) which can, in principle, be designed for any frequency (provided it is lower than the limit frequency of the tube) and which can be tuned as required.

There has naturally been no lack of attempts to raise the limit frequency of triodes by improvements in design. As regards tubes capable of a power output of several hundred watts, we may take as an example the TB 2.5/300 transmitting triode, described in this Review in 1949 [3]), which at 200

*) G. Diemer, K. Rodenhuis and J. G. van Wijngaarden, Philips tech. Rev. **18**, 317-324, 1956/57 (No. 11).

[1]) M. J. O. Strutt and A. van der Ziel, Philips tech. Rev. **3**, 103-111, 1938.

[2]) See, for example, Philips tech. Rev. **14**, 92, 1952/53.

[3]) E. G. Dorgelo, Philips tech. Rev. **10**, 273-281, 1948/49.

Mc/s delivers 200 W with an efficiency of about 60%. In the present article we shall describe some new developments in this field, which have led to the design of the triode TBL 2/300 (*fig. 1*). The limiting frequency of this tube has been raised to



Fig. 1. The transmitting triode TBL 2/300.

900 Mc/s. With a D.C. supply the TBL 2/300 can deliver more than 400 W at 470 Mc/s and about 150 W at 900 Mc/s [4]), with an efficiency of about 63 and 34% respectively. With A.C. supply — which is permissible in diathermy, for example — the power output is about 200 W at 470 Mc/s. Other favourable electrical properties are: good power gain and suitability for wide-band amplification with a relatively low supply voltage (2500 to 1300 V, according to the frequency).

In view of the diverse fields of application of the tube, its mechanical properties have to be taken into consideration. From an electrical point of view it is desirable to have short electrodes, small tube-capacitances, low self-inductance in the leads and short transit times. The first two requirements, calling for electrodes of small dimensions, entail heavy current densities and high specific loading, which adversely affect the tube's useful life. Lower self-inductance calls for leads with a large surface area, e.g. disc electrode leads (see below) and for a planar or a cylindrical electrode system. For short transit times the inter-electrode spacing must be small, and this threatens to conflict with the requirement of mechanical strength. It was therefore necessary to find a compromise. The fact that a favourable compromise was found is due, among other things, to the special construction of the cathode and to the use of a new material for the grid.

Description of the transmitting triode TBL 2/300

The article cited under [3]) gives a number of reasons why, from the electrical point of view as well as from the mechanical and thermal points of view, a cylindrical arrangement of the electrodes in a transmitting tube is preferable to a planar arrangement; among other things, it allows the use of a helical cathode and allows easy alignment of the electrodes. As may be seen from the cross-section in *fig. 2* and from the exploded view in *fig. 3*, a cylindrical arrangement has been adopted for the TBL 2/300. The directly heated cathode, marked *7*

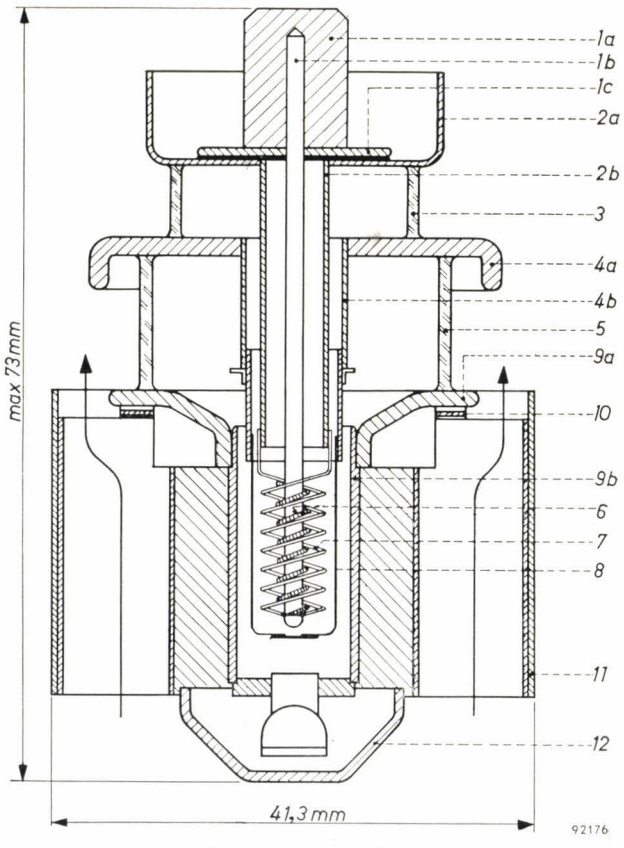


Fig. 2. Cross-section of the tube. *1a*, *2a* filament current connections. *1b*, *2b* mounting pieces for filaments *7* (two coils of thoriated tungsten wire in parallel). *1c-2a* shunt capacitor (sandwich seal). *3*, *5* glass insulating rings. *4a* grid disc. *4b* grid mounting tube. *6* getter (zirconium wire). *8* cage-type grid. *9a* anode disc. *9b* anode. *10* corrugated metal washer. *11* jacket around cooling fins. *12* protective cap.

[4]) For an improved version, still in course of development, which is capable of a somewhat higher power output, see the end of this article.

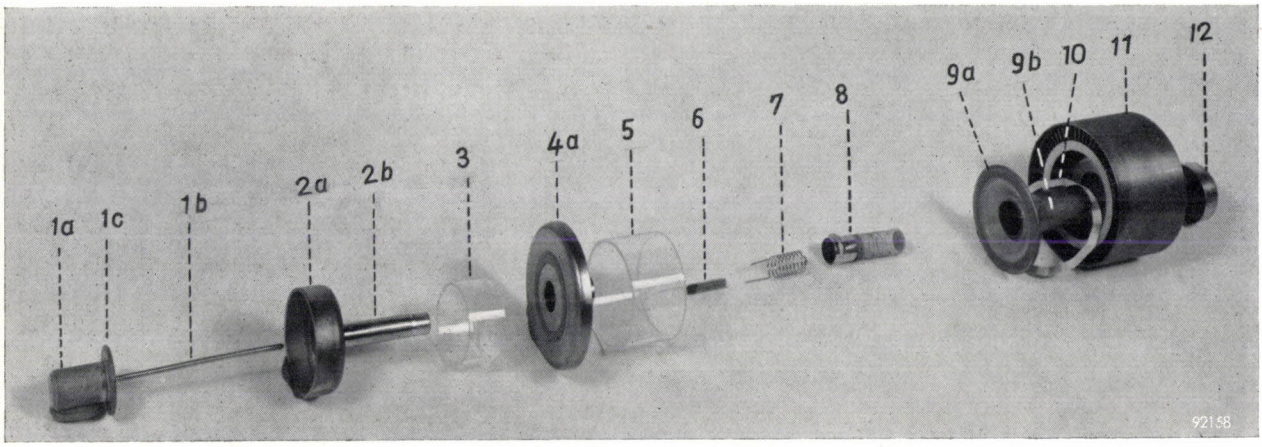
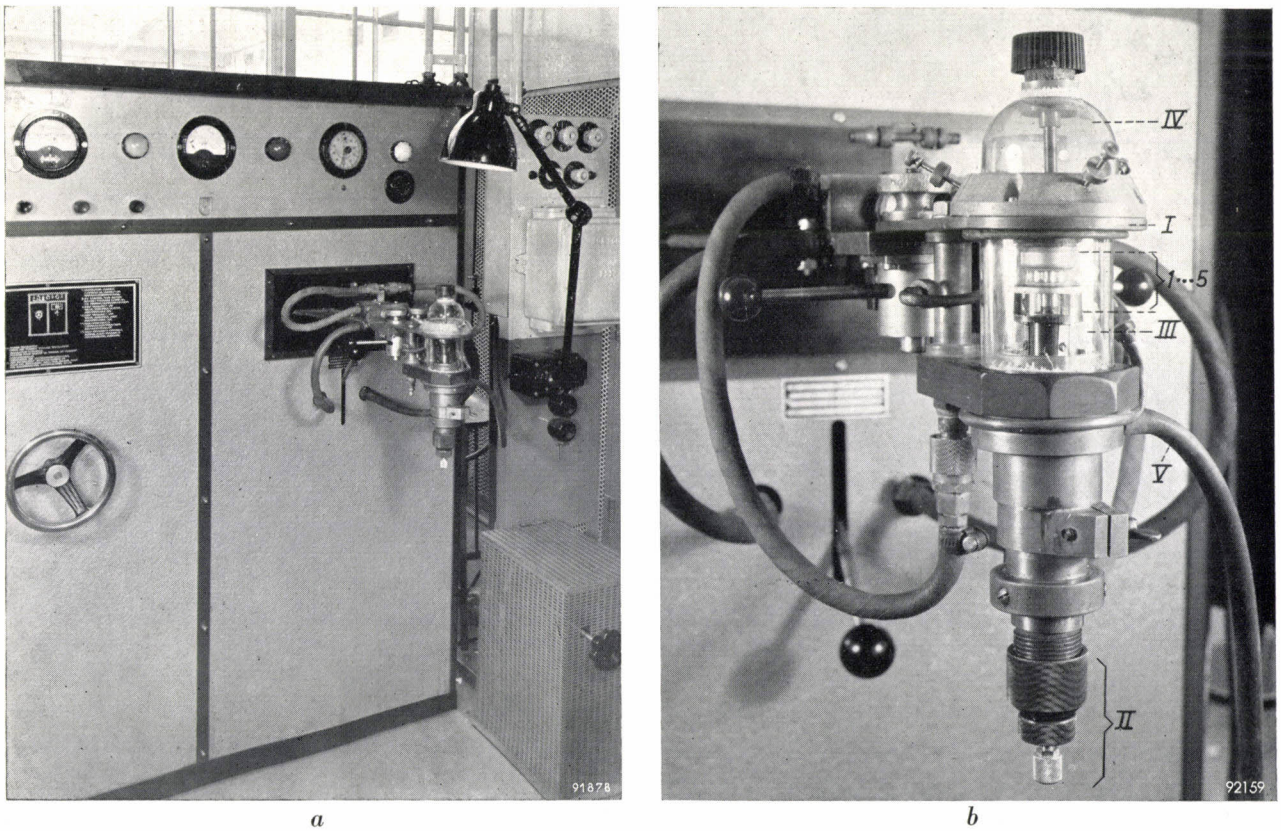


Fig. 3. Exploded view of the TBL 2/300 tube. Symbols as in fig. 2.

in both figures, consists of 2 parallel helices of thoriated tungsten (of which more later). Around the cathode are assembled the cage-type grid *8* and the anode *9b*. The metal parts *2a*, *4a* and *9a* are commonly called discs (although their shape is rather more intricate) and hence the tube is referred to as a disc-seal triode. This design allows the tube to be connected to an external coaxial system. The filament connections consist of the central contact *1a* and disc *2a*, which are connected to the filament via rod *1b* and bush *2b* respectively, both of molybdenum. The rod supports the filament and is partially wound with zirconium wire *6*, which functions as getter. The discs *1c* and *2a*, with a layer of glass between (sandwich seal) constitute a capacitor, the purpose of which will be described later.

Disc *4a* serves as the grid connection. It is



a *b*

Fig. 4. *a*) High-frequency induction heater, used for sealing together the components of the TBL 2/300.
b) The sealing of components *1-5* (see fig. 3). *I* applicator coil, flat in order to concentrate the heating. *II* adjusting screws, for adjusting the height of the components. *III* glass envelope with glass cap *IV*, containing protective gas fed in by hose *V*.

insulated from discs *2a* and *9a* by rings *3* and *5*, which are of hard glass with a high melting point. The discs are of fernico — an alloy of iron, nickel and cobalt, whose coefficient of expansion is close to that of the glass. All external metal parts are silver-plated, to reduce skin-effect losses and also to ensure good contact with the external circuits.

The anode itself (*9b*) is also of fernico. Heat is dissipated via a thick-walled copper cylinder to a large number of cooling fins surrounded by a copper jacket *11*. The cooling surface is 380 cm², and the maximum dissipation is 380 W (300 W anode dissipation, 15 W grid dissipation and 65 W from the filament). Air is blown in through the cooling fins preferably in the direction of the arrows shown in fig. 2, since in that case, owing to the upper extension of the jacket *11*, the emergent air can also pass over and cool the glass ring *5*. If the incoming air is no warmer than 45 °C, an air flow of 0.45 m³ per minute is sufficient, the pressure drop across the assembly then being 24 mm water. The air current required for two tubes can be provided by a small centrifugal fan, driven by a 70 W motor. For a total dissipation of 380 W per tube the temperature of the air passing the cooling assembly rises by about 30 °C.

The jacket *11* also serves as the electrical connection for the anode. The corrugated metal washer *10* ensures good electrical contact via disc *9a*.

The exhaust stem is located at the base of the anode, and is protected by a cap *12*.

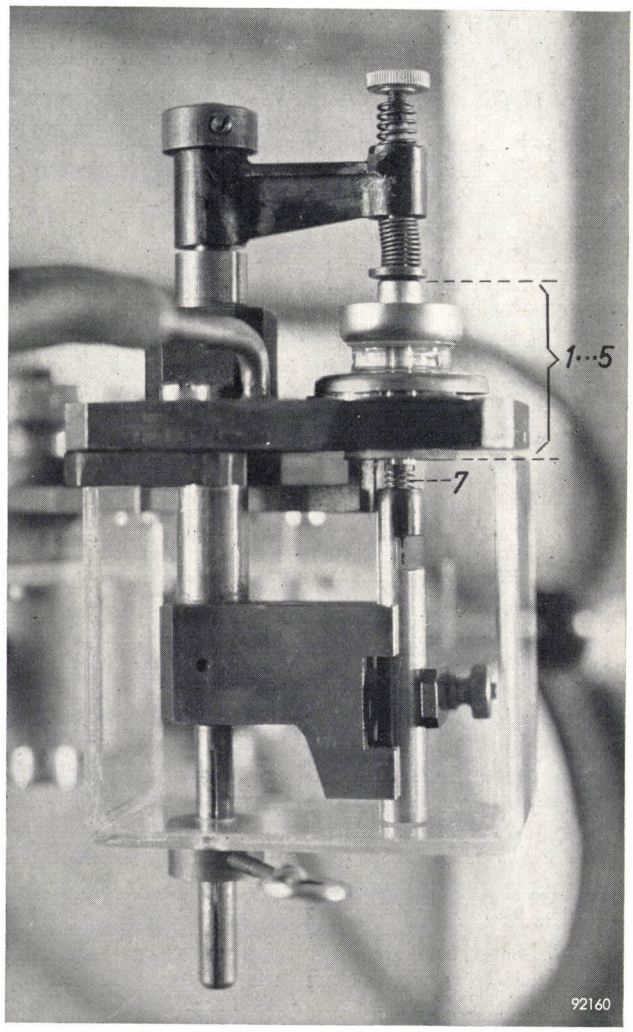


Fig. 5. The filaments *7* are brazed to the assembly *1-5* by high-frequency heating.

Manufacture

Parts *1*, *2*, *3*, *4* and *5* (fig. 3) are placed in a jig and sealed together by high-frequency heating (*fig. 4a*). A flat applicator coil *I* (fig. 4*b*) concentrates the heating in the fernico components, the height of which is accurately adjusted by means of screws *II*. The glass is softened by the radiation from the hot fernico components. All this is done in a protective gas atmosphere, for which purpose the jig is mounted in a glass cylinder *III* with a glass cap *IV*. To the assembly so produced (*1 . . 5*) the cathode filaments are now connected. *Fig. 5* shows how the filaments are brazed to bush *2b* (cf. figs. 2 and 3) by concentrated high-frequency heating. This is also done with the aid of a jig and in a protective gas. To improve their emission properties, the thoriated tungsten filaments are subsequently "carbonized", that is to say, annealed in a carbonaceous gas in order to reduce the thorium oxide. Finally, the getter is introduced.

The grid is accurately brazed concentric to the assembly by a method similar to that used for connecting the filaments. The next process is the sealing-in of the anode, the protective gas now being present only in the tube. After exhausting the tube and cutting off the stem, the cooling-fin assembly is soldered on and the external metal parts silver-plated.

The cathode

As stated, the cathode is of thoriated tungsten. In the article cited under [3]) it was explained that two objections had for a long time precluded the use of this material. The first was grid emission, caused by the thorium evaporating from the cathode and forming a deposit on the grid, and the second was the fact that tantalum, which is costly, was the only metal suitable to be used for the anode, no other metals being known with a sufficiently low gas emission to prevent poisoning of the cathode by traces of oxygen. It was not until the discovery of

the gas-absorbing properties of zirconium [5]) that the thoriated cathode could be used on a wider scale in conjunction with an anode of less costly material; graphite, for example, was sometimes used [3]). As we have seen, the anode in the TBL 2/300 is of fernico. In telecommunications, as in industrial applications, it has been found that the thoriated tungsten cathode, with zirconium as getter, has a long useful life. We shall deal with the question of grid emission presently.

In the design of a cathode for a tube required to operate at frequencies of hundreds of Mc/s, certain other difficulties arise. The first is that, owing to the transit time, the emitted electrons do not pass through the grid in exactly the right phase and some of them therefore return to the cathode. The electron bombardment so produced causes extra heating of the cathode, which can be very damaging in that it gives rise to atomization of the oxide layer. Another reason for the over-heating of the oxide cathode during emission is the poor electrical conductivity at the boundary plane between the nickel and the oxide layer. This does not occur with the thoriated tungsten cathode; moreover, this type of cathode appears to be much better able to withstand electron bombardment, even at current densities from 1 to 1.5 A/cm². A familiar method of reducing the influence of electron bombardment, at frequencies where the transit-time effect is noticeable, is to make the filament voltage lower than at lower frequencies. In the TBL 2/300 the filament voltage need only be slightly lowered. In wide frequency ranges it may even remain constant. The following values are specified:

Frequency	Filament voltage
≦ 600 Mc/s	3.4 V
600-750 Mc/s	3.3 V
750-900 Mc/s	3.2 V

The second difficulty is connected with the requirement that the cathode — where no special measures are adopted — must be very short compared with the shortest wavelength λ_{min} at which the tube is required to operate. To illustrate this, *fig. 6* shows a schematic cross-section of the tube with its associated circuits. Coaxial resonant systems are connected to the cathode K, the grid C and the anode A; the tube is tuned by shifting the shorting-pistons p-p and q-q. We shall now consider the situation in which the distances of p and q to the end m of the diode proper are both $\frac{1}{4}\lambda_{min}$ (approx. 8 cm at 900 Mc/s).

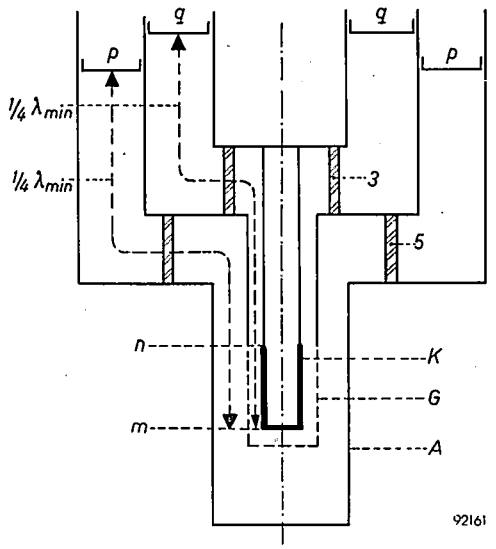


Fig. 6. Schematic cross-section of TBL 2/300 and coaxial resonant system. K cathode. G grid. A anode. *3*, *5* glass rings. p-p annular shorting-piston for anode circuit. q-q annular shorting-piston for grid circuit. The triode proper lies between m and n. The distances mp and mq are $\lambda/4$.

Fig. 7 shows how the high-frequency anode-grid voltage V_{ag} and cathode-grid voltage V_{kg} vary from the end m to the shorting-pistons p and q. In fig. 7*a* it is assumed that the cathode is a small tube (directly or indirectly heated). The length mn of the triode proper is about $\frac{1}{4}$ of the length $mp = mq = \frac{1}{4}\lambda_{min}$, and consequently the alternating potential differences V_{ag} and V_{kg} along the anode and cathode respectively are small. However, a thin-walled tube, as would be needed for a cathode, cannot be drawn from thoriated tungsten. In practice, therefore, we are obliged to adopt a (directly heated) wire cathode. If a very high filament current is to be avoided, the wire must be fairly thin and long. For this reason a helical form is to be preferred (fig. 7*b*). In this case, however, ultra-high frequency operation is impossible because if the length of the filament is about $\frac{1}{2}\lambda_{min}$, a voltage node will occur in the middle of the cathode. At that point V_{kg} will be zero and thus there will be no net anode alternating current, while on either side of the node the anode currents will be in anti-phase; the total anode alternating current will then be almost zero and the tube will be unable to operate.

[5]) J. H. de Boer and J. D. Fast, Rec. Trav. chim. Pays-Bas **55**, 459-467, 1936; J. D. Fast, Philips tech. Rev. **5**, 217-221, 1940.

The solution of this difficulty is to take two helices, connected in parallel and each with a length of 7.5 cm ($\approx \frac{1}{4}\lambda_{min}$), and to shunt them with a capacitor C of such a capacitance that the end m (fig. 7c) is at almost the same high-frequency potential as the end n. The result is that V_{kg} is practically constant over the entire length of the cathode, so that the cathode at high frequencies is almost an equipotential surface. The form of V_{kg} shown in fig. 7c was measured on an enlarged model (designed by J. M. van Hofweegen) with a correspondingly larger wave length.

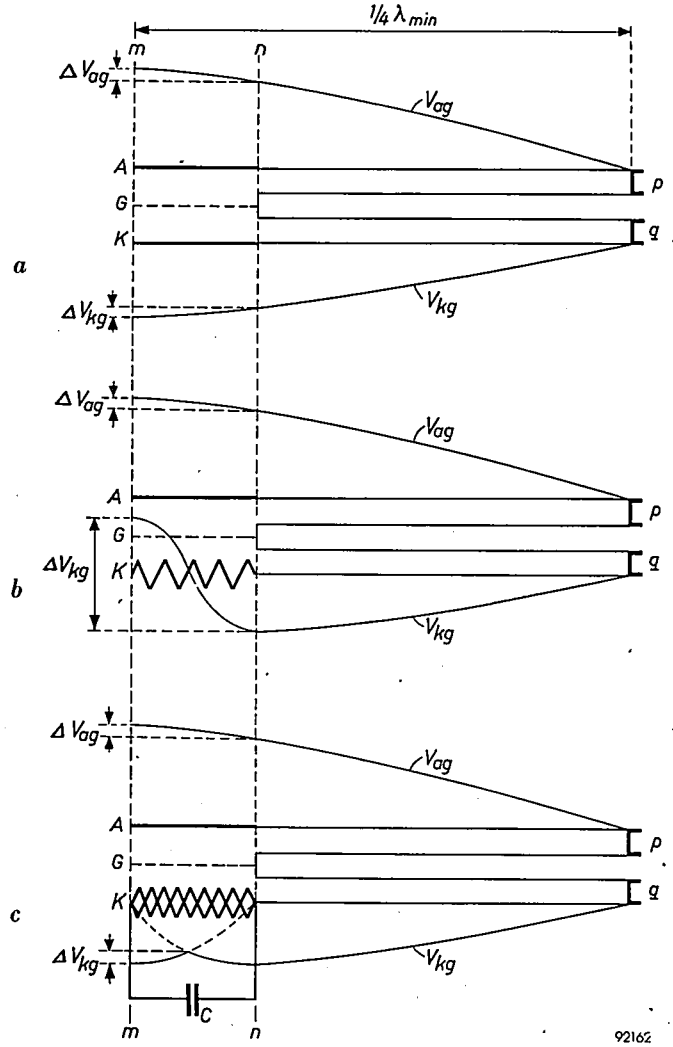


Fig. 7. K cathode, G grid and A anode of triode (schematic) showing the variation of the alternating voltages V_{ag} between anode and grid, and V_{kg} between cathode and grid, along the length of the anode and cathode respectively. ΔV_{ag} and ΔV_{kg} potential differences between the ends of these electrodes. p, q shorting-pistons (cf. fig. 6). $mp = mq = \frac{1}{4}\lambda_{min}$.
a) Cylindrical cathode. ΔV_{ag} and ΔV_{kg} are small if the electrodes are short compared with $\frac{1}{4}\lambda$. The cathode is then an almost equipotential surface.
b) Helical wire cathode, with wire length $\frac{1}{2}\lambda$. The two halves are in anti-phase and therefore work in opposition.
c) The cathode consists of two coils, each $\lambda/4$ in length, connected in parallel and shunted by the capacitor C. The potential difference ΔV_{kg} is about as small as in (*a*).

Fig. 8 shows the connections to the cathode (the two helical filaments are drawn side by side for clarity). Capacitor C is formed by the sandwich seal already mentioned (fig. 2), the capacitance of which is about 50 pF. At a filament voltage of 3.4 V the filament current is about 19 A.

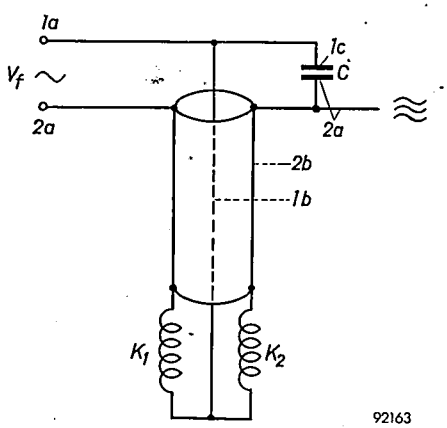


Fig. 8. The two coiled filaments K_1 and K_2 (shown side by side for clarity) and the capacitor C. For *1a*, *1b*, *1c*, *2a* and *2b*, see figs. 2 and 3. The filament voltage V_f is applied between *1a* and *2a*. *2a* is also the connection for the high-frequency current.

The grid

The cage-type grid consists of a large number of thin rods interconnected by 5 rings. The grid thus adequately approximates electrically to a conducting, entirely enclosed, surface, which is necessary in order to produce the V_{kg} distribution shown in fig. 7c.

An old problem encountered with transmitting tubes is grid emission. Substances evaporating from the cathode (in this case thorium) can, if they settle on the grid, cause the grid to emit electrons quite profusely. These electrons constitute a current in opposition to the normal grid current. If the negative grid voltage is produced by means of a leak resistor, the drop in the total grid current can result in instability. The risk of instability is particularly great if the tube is suddenly fully loaded after the filament has been switched on for a long period without any potential on the anode: during such a period a great deal of cathode dust will have settled on the grid, and when the load is applied the grid becomes hot and hence strongly emissive.

Several substances, such as zirconium oxide and platinum are known to have a low emission when used as grid material. These substances have been thoroughly tested in experimental TBL 2/300 tubes, both in new tubes and in tubes that had already operated 1000 hours with the high specific grid load of 15 W/cm². The results are set out in *fig. 9*, in which the specific grid emission current is logarith-

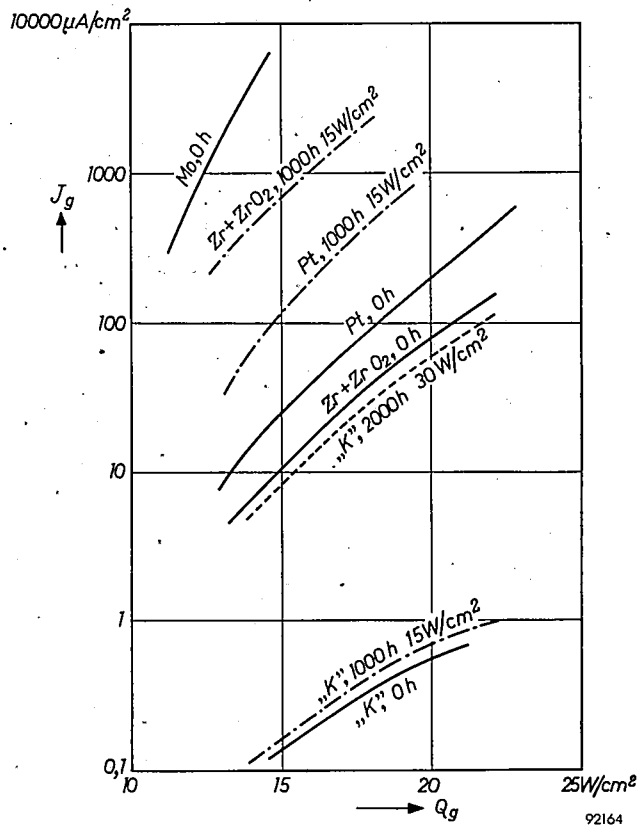


Fig. 9. Specific grid emission J_g (on a logarithmic scale) of experimental TBL 2/300 tubes, with grids of different materials as a function of specific grid dissipation Q_g. The fully-drawn curves apply to new tubes. The dot-dash curves are measurements made after 1000 hours at $Q_g = 15$ W/cm². The dashed curve refers to K material, measured after 2000 hours at $Q_g = 30$ W/cm². *Mo* = molybdenum. *Pt* = platinum on molybdenum. $ZrO_2 + Zr$ = zirconium oxide and zirconium on molybdenum. "K" = K material.

mically plotted as a function of specific grid load. It can be seen that of all materials tested, molybdenum is the least suitable, having by far the strongest emission. Platinum, when fresh (plated on a molybdenum core), is appreciably better, and zirconium oxide (mixed with zirconium and also coated on Mo) is slightly better still. After 1000 hours at 15 W/cm², however, the emission from the latter material has become greater than that of Pt and approaches fairly closely the emission of uncoated Mo.

A substantial improvement is obtained with a new material, referred to as "K material". Its emission is about 1/100 of that of $ZrO_2 + Zr$ or of Pt when fresh, and after 1000 hours at 15 W/cm² it shows hardly any increase. Not until after 2000 hours at 30 W/cm² (four times higher than the specified load) does the emission approach that of fresh $ZrO_2 + Zr$. Owing to its great ductility, the use of K material for the grids makes the tube well able to withstand shocks and vibrations. This is an important point, having regard to the uses of the TBL 2/300 in mobile equipment and in industry. This favourable combination of electrical and mechanical properties prompted the use of K material for the grid of the TBL 2/300. The geometry of the electrodes is such that the specific grid dissipation during normal operation lies far below 15 W/cm². The ultimate limiting power loading of the grid, i.e. the loading at which the grid is immediately destroyed, lies at about 60 W/cm². The wide margin between this limiting value and the normal loading of the grid is

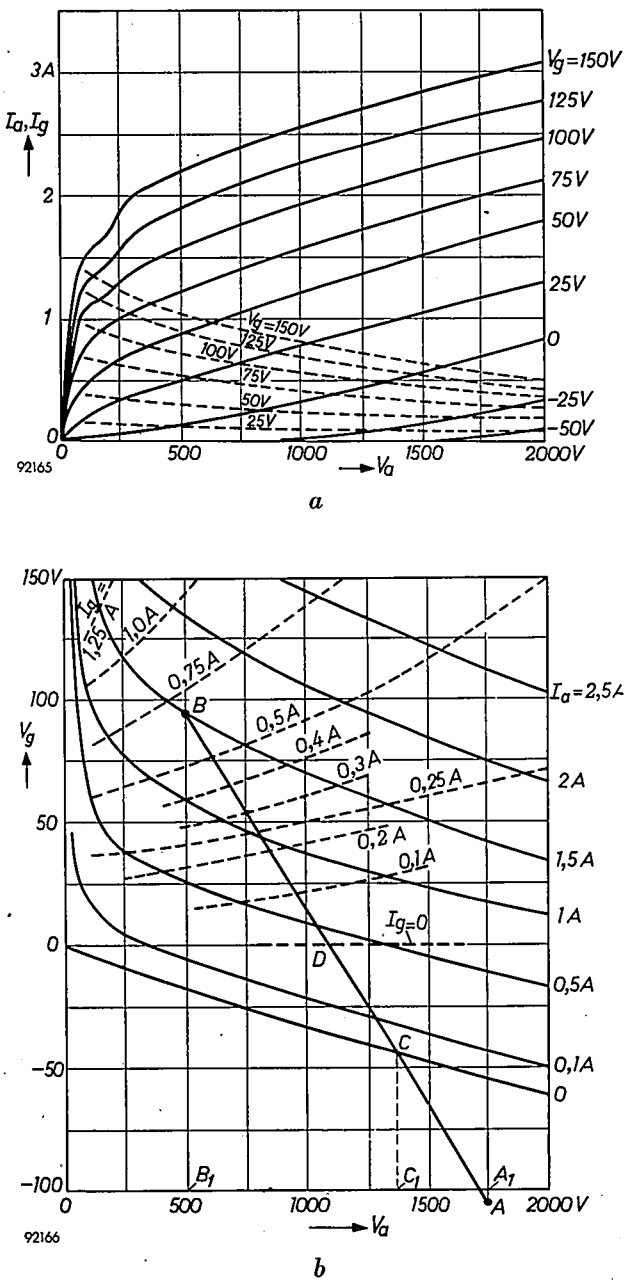


Fig. 10. Static characteristics of the TBL 2/300.
a) Anode current I_a (full) and grid current I_g (dashed) as a function of anode voltage V_a, with grid voltage V_g as the running parameter.
b) V_g as a function of V_a for constant I_a (full) and for constant I_g (dashed).

an important practical advantage in the adjustment of an oscillator in course of construction or development.

Electrical properties

Static characteristics

In *fig. 10a* the anode current I_a and the grid current I_g are plotted as a function of the anode voltage V_a, with the grid voltage V_g as the running parameter. Fig. 10*b*, which is derived from fig. 10*a*, represents V_g as a function of V_a; the fully-drawn curves hold for constant I_a, the dashed curves for constant I_g. This diagram has the advantage that the working line is straight, which simplifies the calculations for the biassing of the tube.

As may be derived from fig. 10*a*, the slope S is approximately 18 mA/V and the amplification factor μ about 32.

Use in oscillator circuits

Extensive measurements and life tests (several thousands of hours) have been carried out on many tubes of type TBL 2/300 in different oscillator circuits. One of these oscillators operates at 840 Mc/s: some constructional details are given in *fig. 11*, the actual arrangement is shown in *fig. 12*. In *fig. 13* the frequency of this oscillator is plotted as a function of the height H of the interchangeable anode cavity resonator, for diameters $D = 90$ and 180 mm. Furthermore, TBL 2/300 tubes are being used in an experimental television transmitter which has been in operation at Eindhoven for some considerable time (picture 772.25 Mc/s, sound 777.75 Mc/s).

The tests have shown that the H.T. supply may permissibly be 2500 V at frequencies up to 200 Mc/s,

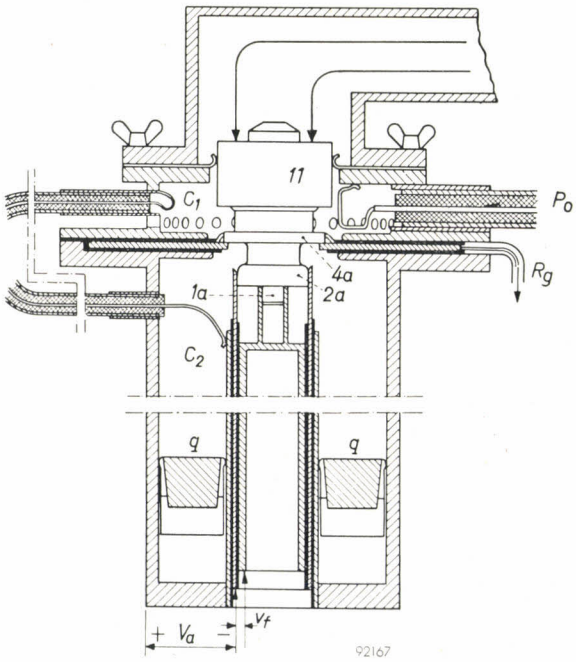


Fig. 11. Section through the coaxial system of an oscillator used for testing TBL 2/300 tubes at 840 Mc/s. *1a*, *2a* filament contacts, *4a* grid disc and *11* anode disc. C_1 interchangeable anode cavity resonator. C_2 grid cavity resonator, tuned by shorting-piston *q*. *Left*: the coaxial cable which provides for feed-back between the two cavity resonators. *Right*: the cable which takes off the output power P_o, and the connection for the grid resistor R_g. Connections for filament voltage V_f and the H.T. supply V_a are brought in from below. Air is blown in through the top.

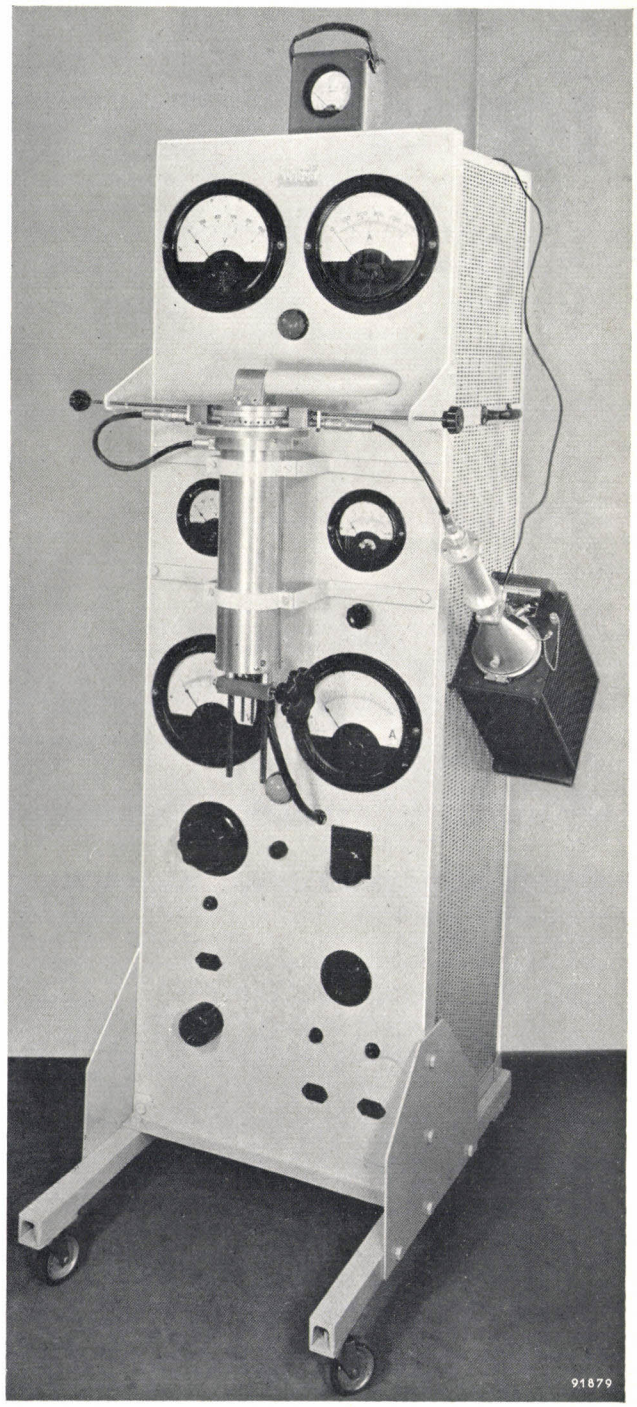


Fig. 12. Test oscillator operating at 840 Mc/s. The tube TBL 2/300 is mounted in the metal cylinder in the foreground; the load is shown on the right.

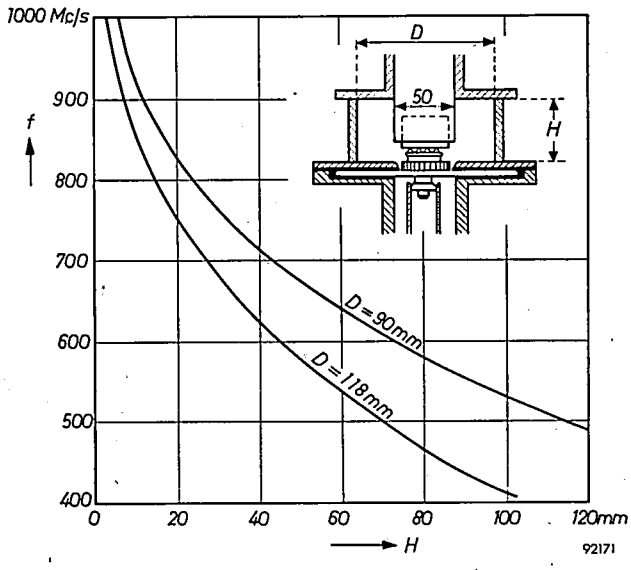


Fig. 13. The frequency f as a function of the height H of the anode cavity resonator, for two different diameters D. Inset: schematic cross-section of cavity resonator (cf. fig. 11).

but that it must be lowered according as the frequency rises (to 1300 V at 900 Mc/s) in view of dielectric losses in the glass. Partly as a result of this, the maximum power output P_0 falls from 475 W at 200 Mc/s to some 150 W at 900 Mc/s. *Table I* gives the values measured for normal operation and the maximum values; see also *fig. 14*.

With regard to the grid dissipation, it should be added that apart from the dissipation P_g of about 6 W caused by electron bombardment, heat is also generated in the grid rods owing to the skin effect. The contribution which this makes to the total grid dissipation is, in proportion, quite considerable, being about 6 W at 470 Mc/s and about 8 W at 900 Mc/s. The grid may thus have altogether about $6 + 8 = 14$ W to dissipate. The grid surface area being 2 cm², the total specific grid load is 7 W/cm², which is far below the specific load of 15 W/cm² at which the K material has been life-tested, and even further below the ultimate limit of 60 W/cm². (These values do not include the (constant) radiation heating of the grid by the filament.)

As already remarked, it is often permissible in diathermy (and also in industrial high-frequency heating) to operate the tube on alternating voltage. This saves the costs and electrical losses entailed by the use of a rectifier and a smoothing filter. Fed by alternating voltage of 1900 V r.m.s. and a mean anode current of 166 mA, the tube consumes a power of $(\pi/2\sqrt{2}) \times 1900 \times 0.166 = 350$ W; at a frequency of 460 Mc/s, the power output is then 227 W, the efficiency being 65%. Of this output a part is lost in the tuned circuit. That part can amount

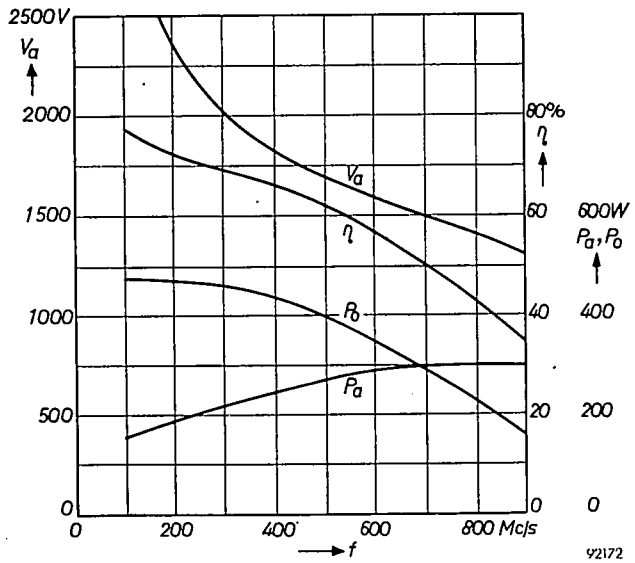


Fig. 14. Maximum permissible H.T. voltage V_a, maximum permissible dissipation P_a, power output P_0 and efficiency η of an oscillating TBL 2/300 tube, as a function of frequency f.

Table I. Voltages, currents, powers and efficiency of TBL 2/300 tubes in an oscillator circuit. The values between brackets are the maximum permissible values.

Frequency (Mc/s)		175		470		900	
Wavelength (cm)		172		64		33	
Filament voltage	(V)	3.4		3.4		3.2	
Anode supply voltage	(V)	2500	(2500)	1750	(1750)	1300	(1300)
Anode direct current	(mA)	260	(400)	380	(400)	350	(400)
Grid D.C. voltage	(V)	—200	(—300)	—105	(—300)	—60	(—300)
Grid A.C. voltage (peak)	(V)	275		190		—	
Grid direct current	(mA)	100	(120)	100	(120)	100	(120)
Input power P_i	(W)	650	(1000)	665	(700)	455	(520)
Anode dissipation P_a	(W)	175	(300)	260	(300)	300	(300)
Grid dissipation P_g	(W)	~6	(15)	~6	(15)	~6	(15)
Power output P_0	(W)	475		405		155	
Efficiency P_0/P_i	(%)	73		61		34	

to about $\frac{1}{3}$, so that about 150 W useful output remains. In diathermy and heating applications, the load is not usually continuous. The tube, which goes on oscillating when the load is removed, then takes a high grid current, which increases from, say, 65 to 95 mA, while the supply voltage increases from 1900 to 1950 V. Even in these conditions, however, the tube life was found to be satisfactory.

Power gain and bandwidth

The power gain obtainable with the TBL 2/300 at 470 Mc/s is about 15 with grounded cathode and about 5 with grounded grid.

The tube capacitances are:

Anode-grid:	C_{ag}	= approx. 4	pF
Anode-cathode:	C_{ak}	= approx. 0.12	pF
Grid-cathode:	C_{gk}	= approx. 9	pF

For wide-band amplification it is important that C_{ag} should be small, as follows from the expression for the bandwidth B:

$$B = \frac{1}{2\pi R_a C_{ag}}.$$

In this expression R_a is the load resistance (the ratio of the alternating voltage of the anode to the fundamental of the anode current). With C_{ag} = 4 pF and R_a = 2000 Ω, the bandwidth is 20 Mc/s, which is more than adequate for all existing television systems and can also accommodate a large number of telegraphy and telephony channels.

For the power gain G_k in the grounded cathode arrangement we may write:

$$G_k = \frac{P_o}{P_{ig}} = \frac{\frac{1}{2}V_{a1}I_{a1}}{\frac{1}{2}V_{g1}I_{g1}}.$$

P_{ig} is here the input power to be amplified, applied between grid and cathode, V_{a1}, I_{a1}, V_{g1} and I_{g1} represent the peak values of the fundamentals of anode voltage, anode current, grid voltage and grid current, respectively.

By way of example we shall take a case in which the frequency is 470 Mc/s and the maximum permissible anode supply voltage 1750 V. Half of the selected load line (AB) is shown in fig. 10*b*. The centre A of the load line lies at V_a = 1750 V, V_g = −105 V; end B lies at $I_{a\,max}$ = 1.5 A. The maximum grid current $I_{g\,max}$ is 0.7 A, the minimum anode voltage $V_{a\,min}$ is 500 V. The amplitude V_{a1} of the anode alternating voltage is thus 1750 − 500 = 1250 V.

The amplitude I_{a1} of the fundamental of the anode alternating current may be found by first determining the "conduction angle" $2\Theta_a$: the part of the period in which anode current flows, i.e. when the tube is conducting, is $2\Theta_a/2\pi$ (*fig. 15a*). It follows from fig. 15*a* and *b* that:

$$\cos\Theta_a = \frac{A_1C_1}{A_1B_1}.$$

According to fig. 10*b* this is equal to AC/AB, the value of which can be found from this figure to be

$$\cos\Theta_a = \frac{AC}{AB} = 0.325,$$

so that

$$\Theta_a = 72°.$$

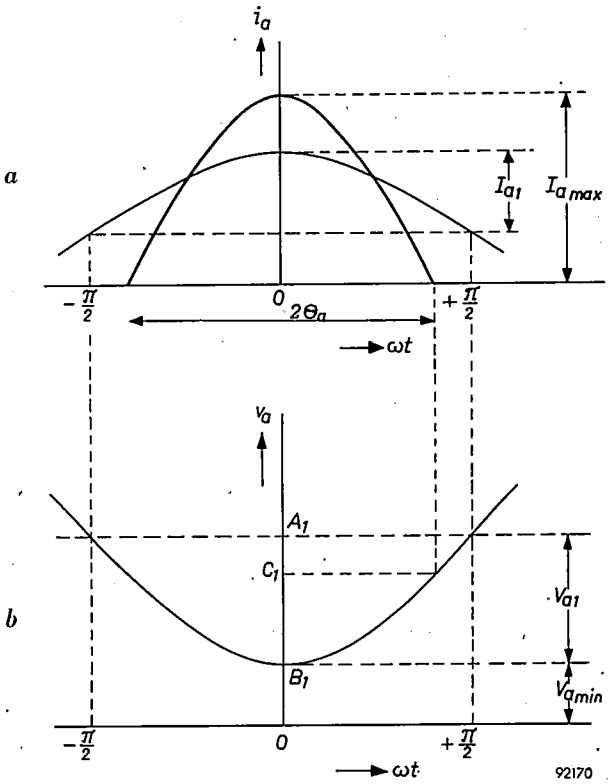


Fig. 15. *a*) Anode current i_a, (*b*) anode voltage v_a of an oscillating tube, as functions of ωt (ω = angular frequency, t = time). $2\Theta_a$ = conduction angle. Thin curve in (*a*): the sum of the fundamental of i_a (amplitude I_{a1}) and the D.C. component.

From the conduction angle we can derive the ratio $I_{a1}/I_{a\,max}$ [6]); for Θ_a = 72°, the value of $I_{a1}/I_{a\,max}$ is 0.43, so that

$$I_{a1} = 0.43 \times 1.5 = 0.65 \text{ A}.$$

With this we find for the anode resistance $R_a = V_{a1}/I_{a1}$ = 1250/0.65 = 1930 Ω (in the foregoing this was rounded off to 2000 Ω), and for the output power P_o:

$$P_o = \tfrac{1}{2} \times 1250 \times 0.65 = 405 \text{ W}.$$

V_{g1} (the difference between the ordinates of B and A in fig. 10*b*) is seen to be 200 V. For half the conduction angle Θ_g of the grid current we may write:

$$\cos\Theta_g = \frac{AD}{AB} = 0.52,$$

hence

$$\Theta_g = 58°.$$

[6]) J. P. Heyboer, Transmitting Valves, Philips Technical Library 1951, fig. 19, p. 38.

The corresponding value of $I_{g1}/I_{g\,max}$ is

$$\frac{I_{g1}}{I_{g\,max}} = 0.38,$$

so that

$$I_{g1} = 0.38 \times 0.70 = 0.27 \text{ A}$$

and

$$P_{ig} = \tfrac{1}{2} \times 200 \times 0.27 = 27 \text{ W}.$$

The power gain is therefore:

$$G_k = \frac{405}{27} = 15.$$

In grounded-grid arrangement the power gain G_g is:

$$G_g = \frac{P_o + P_d}{P_{ig} + P_d},$$

in which P_d is the power directly transferred from input to output, equal to $\frac{1}{2} V_{g1} I_{a1}$. In the above example $P_d = 65$ W. Thus

$$G_g = \frac{405+65}{27+65} = 5.1.$$

Ceramic version of the tube

Finally, a few words about the latest version of the TBL 2/300, which is still in course of development. In this version the glass rings *3* and *5* (see figs. 2 and 3) have been replaced by ceramic rings. The ceramic material has an appreciably lower loss factor tan δ than the best types of hard glass suitable for fusion to fernico. Consequently, the supply voltage does not have to be lowered so much as the frequency rises, which means that the tube has a higher power output at the highest frequencies.

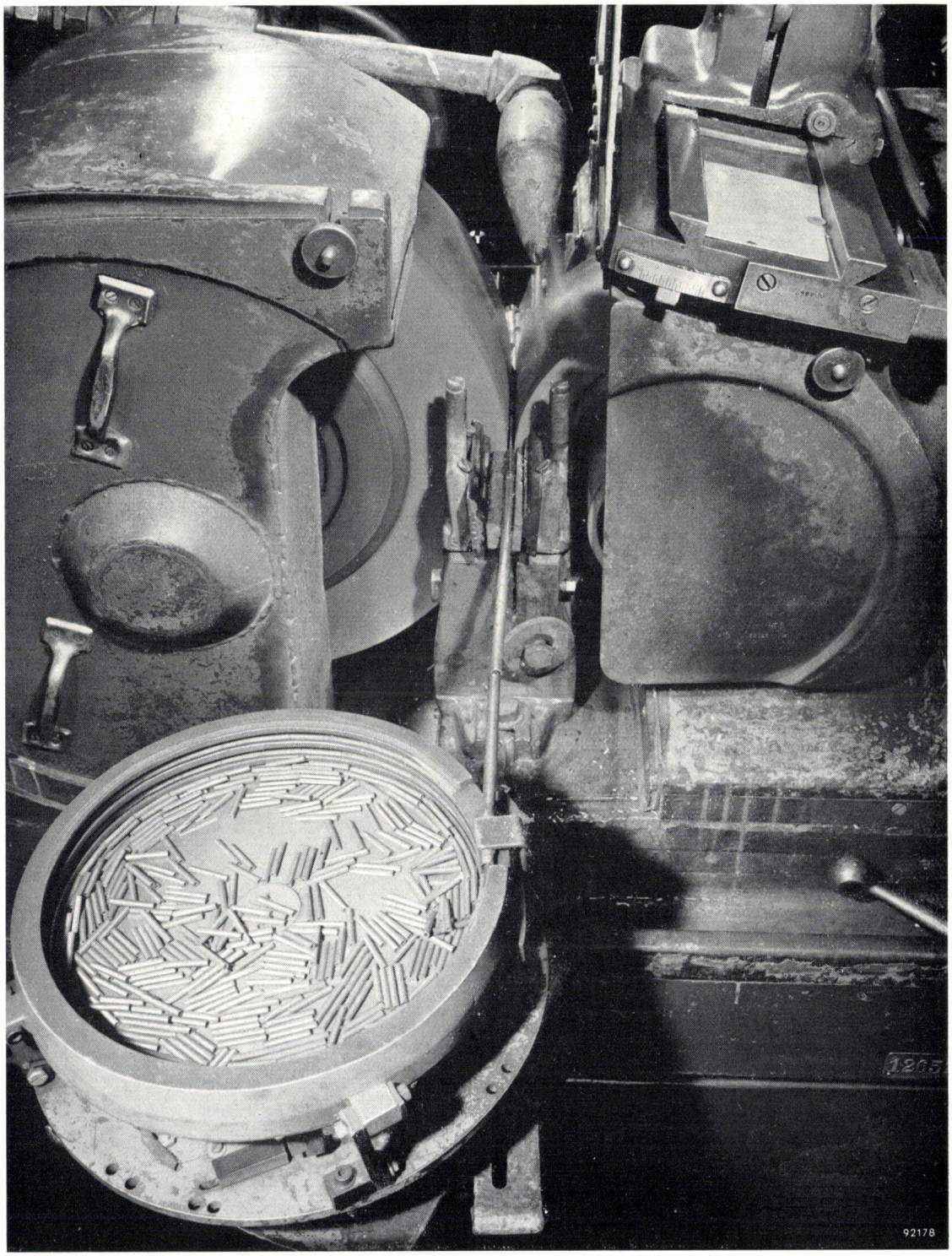
The ceramic-metal seal was an initial difficulty. A satisfactory ceramic-metal seal evolved by Radioröhrenfabrik Valvo in Hamburg will be described in a subsequent article in this Review.

Summary. The TBL 2/300 transmitting tube for frequencies up to 900 Mc/s is a disc-seal triode with a directly heated cathode of thoriated tungsten, a cage-type grid and an air-cooled anode. The cathode consists of two helical filaments connected in parallel, each with a wire length of about $^1/_4$ of the shortest wavelength and shunted by a capacitor, the capacitance of which (50 pF) is such that at high frequencies the cathode is almost equipotential. The capacitor is formed by a so-called sandwich seal. The grid is made of a new material ("K material"), which has extremely low emission properties, even after the tube has operated for a long period with high grid dissipation; this material also has good mechanical strength.

The slope of the tube is about 18 mA/V, the amplification factor about 32. In oscillator circuits it delivers a power of 405 W at 470 Mc/s, and 155 W at 900 Mc/s, with an efficiency of 61% and 34% respectively. The power gain is about 15 with grounded cathode and about 5 with grounded grid (both at 470 Mc/s). Owing to its low grid-anode capacitance (about 4 pF), the tube still amplifies efficiently at bandwidths up to 20 Mc/s.

A new version in course of development, in which a ceramic is used instead of glass for the insulating rings, has a somewhat higher output.

CENTRELESS GRINDING OF CERAMIC SLEEVES



Ceramic sleeves of various types and sizes are used on a large scale in the electronics industry. The photograph shows such sleeves undergoing centreless grinding down to a certain external diameter. A circular vibratory hopper (foreground) feeds the sleeves via a metal tube to the grinding wheel and pressure roller.

A TRANSISTOR HEARING AID

by P. BLOM and P. BOXMAN.

621.395.92:621.375.4

Some time ago a short report appeared in this Review on the transistor hearing-aid type KL 5500. A more detailed description of this apparatus illustrates clearly the great advantages resulting from the use of transistors.

The entry of the transistor into electronic engineering has opened up entirely new possibilities. The most outstanding advantages of the transistor over thermionic valves are its small dimensions and low power consumption, the latter mainly because no filament current is required. Moreover, the power can be supplied by a source of very low voltage. This is particularly advantageous where the source consists of dry batteries, since the energy (the number of watt hours) is supplied much more cheaply by a battery of a few volts than by a battery of some tens of volts. Furthermore, practice has already shown that, in general, transistors may be expected to have a longer life than valves.

Small dimensions and low battery costs are points of especial importance where hearing aids are concerned. It is not surprising, therefore, that the first practical application of transistors on a large scale has been in hearing aids, particularly since what is required is amplification at audio frequencies and relatively low power outputs, which is precisely the field of application for which transistors were first manufactured on a commercial basis. The transistors in question are nevertheless capable of supplying a higher output power than is economically possible with the subminiature valves used in valve hearing aids.

Transistors offer certain other advantages which make them particularly suitable for hearing aids. In the circuit with a common emitter, which is the circuit mostly employed at low frequencies, the impedances involved are relatively low [1]). Even in a very compact construction there is therefore nothing to fear from stray capacitance coupling: the stray capacitances form high impedances carrying negligible currents. What is more, transistors show no microphonic effect; the measures normally needed to combat this effect can therefore be dispensed with, which again makes for compactness.

With the Philips transistor hearing aid described here (type KL 5500) [2]), the aim has been to help the largest possible number of users. Accordingly this is still the hearing aid with the widest field of application. Two other types are also being manufactured: type KL 5600, which corresponds substantially to KL 5500 but is smaller (with the sacrifice of some of its possibilities and entailing higher battery costs) and the even smaller type KL 5700, which is in a certain sense a de luxe model, combining minimum dimensions with high quality. In *fig. 1* the three hearing aids are shown side by side.

Requirements to be satisfied by a hearing aid

Speech intelligibility, which is of course the most important criterion for the usefulness of a hearing aid, is a problem which has been solved in modern instruments to the satisfaction of a wide circle of the hard of hearing. Although new measures to improve speech intelligibility are continually being developed — particularly as regards special types of deafness — more and more importance is now being attached to requirements of a more secondary nature concerned with the need to hear comfortably without great effort. For all types of deafness, comfortable hearing is determined in the first place by a favourable signal-to-noise ratio, and in the second place by low non-linear distortion (by non-linear distortion, the instrument itself adds new frequencies to the received signal). Noise in hearing aids is produced partly by the circuit itself and partly by friction between parts of clothing near the apparatus or between clothing and the case of the apparatus (case noise).

The more detailed requirements to be satisfied by a hearing aid depend to a large extent upon the nature of the patient's deafness. A hearing aid capable of meeting the needs of a wide circle of users must therefore possess considerable flexibility.

[1]) For the three principle transistor configurations, viz. the circuits with common emitter, common base and common collector, and for a survey of some of their fundamental properties, see for example J. P. Beijersbergen, M. Beun and J. te Winkel, The junction transistor as a network element at low frequencies, I. Characteristics and h parameters, Philips tech. Rev. **19**, 15-27, 1957/58 (No. 1).

[2]) Briefly described in Philips tech. Rev. **17**, 315, 1955/56.

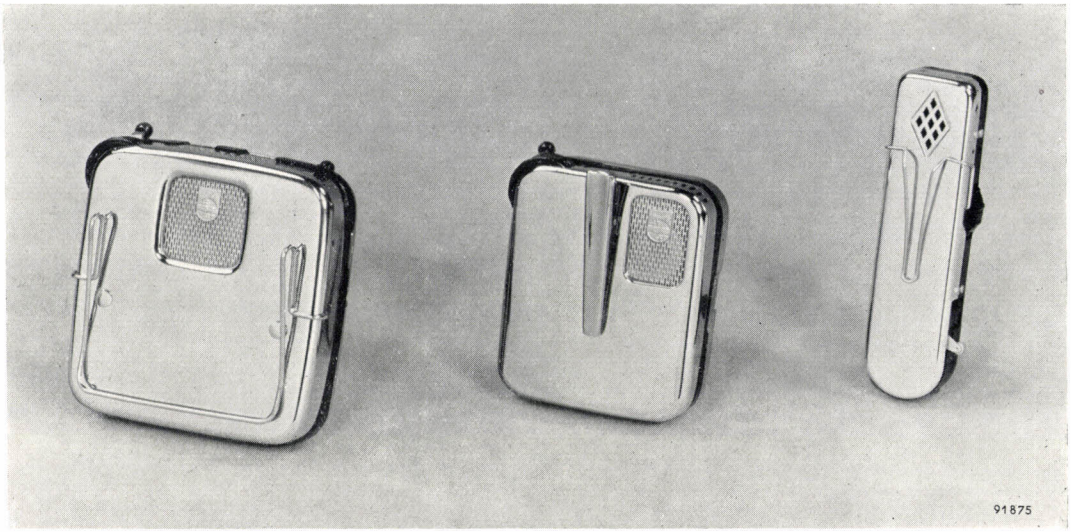


Fig. 1. The three Philips transistor hearing aids, types KL 5500, KL 5600 and KL 5700 (from left to right). The one on the left, which has the widest field of application, is discussed in this article.

In most cases a maximum acoustic gain of 55 dB appears to be more than sufficient. Of this, depending upon the degree of his deafness and upon the intensity of sound in the microphone, the patient can use what he needs by varying a volume control. In cases of severe deafness, however, it is desirable to be able to boost the maximum acoustic gain up to, say, 65 dB.

Furthermore, for the amplification to be effective it should be possible to produce a sufficiently large sound pressure in the ear without this being associated with serious distortion. The earphones commonly used can readily produce the required pressure, provided the output transistor in the hearing aid can supply the power. In most cases, a maximum available output power of one milliwatt is more than enough for driving ordinary earphones. In cases of very severe deafness, however, it may be desirable to increase the output to as much as 10 milliwatts. The same higher output may also be needed when bone-conduction earphones are used. With an earphone of this type, which is fixed behind the ear by means of a headpiece, the sound vibrations are conducted to the organ of hearing via the bones of the skull. For this purpose, more power is needed than for the normal conduction of sound via the air in the auditory passage.

For some patients, particularly for sufferers from conduction deafness, an amplification independent of the amplitude of the input signal is suitable. The amount of the amplification will differ for different patients: where a high amplification is needed, a large maximum power output must, of course, also be available.

There are also patients, however, for whom it would be quite wrong to have an amplification independent of the amplitude of the input signal. Such patients include those who can only hear sounds whose intensity lies above a certain level, but for whom the threshold of pain ("acoustic trauma") is about the same as for persons with normal hearing. This situation is found in the case of sufferers from regression deafness [3]). For such patients the output power of the hearing aid must be prevented from exceeding the value corresponding to the threshold of pain, as the amplitude of the signal in the microphone increases. This means that the amplification and the maximum available power must be separately adjustable.

Apart from the amplification as a function of the amplitude of the received signal, the amplification as a function of the frequency is of considerable importance. The curve illustrating this dependency, is called the response curve, or "fidelity characteristic", of the hearing aid. A flat curve is not usually required. What is aimed at is an agreeable tone quality, but the meaning of "agreeable" in this context depends upon the nature of the hearing defect and upon the opinion of the patient. This does not exclude cases where the patient must, as it were, reassemble his "library" of sounds from a spectrum which, though he may find it troublesome, perhaps, in the beginning, he finally comes to accept as true. In order to be able to meet all these diverse re-

[3]) The properties of the human ear and the various forms of deafness are briefly discussed in an article by P. Blom: An electronic hearing aid, Philips tech. Rev. **15**, 37-48, 1953/54, which describes a valve hearing aid.

quirements, it must be possible to regulate the instrument's response curve.

The hearing aid KL 5500

The amplification and power outputs mentioned above can be attained economically with valve hearing aids only by using batteries of a relatively large volume, and this would make the hearing aid too large to be acceptable nowadays. If the size of such a valve instrument is kept acceptably small by the use of miniature batteries, the battery costs are exorbitant, since small batteries are much less efficient than large ones. The properties of the transistor make it possible to solve the problem. The Philips hearing aid type KL 5500 contains four transistors (three of type OC 70 and one, the output transistor, of type OC 71), all connected in the common-emitter configuration.

The battery holder is designed for a carbon-zinc rod-type battery, consisting of a single cell which supplies a voltage of 1.5 V. The output obtained is more than 1 mW, which is quite sufficient in all except extreme cases of deafness.

Since the cell used is relatively large (*fig. 2*), the battery costs are extremely low, amounting to about

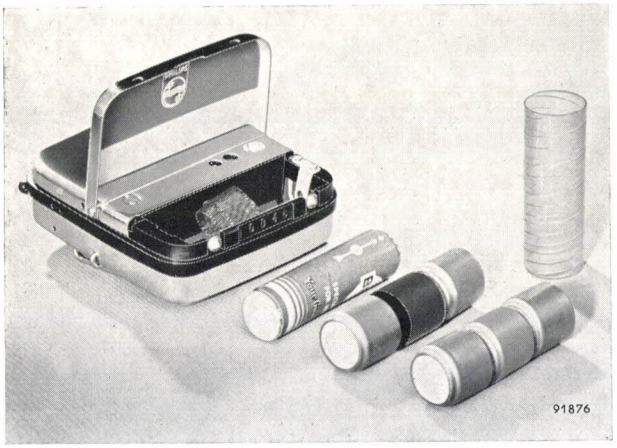


Fig. 2. The battery holder of the KL 5500 hearing aid, with the three types of batteries that may be used: a Leclanché dry cell (1.5 V), two mercury cells with filler piece (2.6 V) or 3 mercury cells (3.9 V).

one tenth of the costs of a comparable valve hearing aid. A larger power output and more amplification become available if the battery voltage is increased, which can be done by using two or three mercury cells of 1.3 V each (fig. 2). The available output power is then 4 and 10 mW respectively. This, however, considerably increases the battery costs. In the first place, current consumption increases in proportion to the voltage; the replacement of the carbon-zinc cell by three mercury cells means, therefore, that the current consumption increases roughly by a factor of 3. Thus, although the number of mA-hours of a mercury cell is about 1.5 times larger than that of a carbon-zinc cell, about six mercury cells will be used up for every one carbon-zinc cell. In the second place a mercury cell costs much more than a carbon-zinc cell (prices differ from one country to another). In order to save battery costs in cases where a supply voltage of about 4 V is needed, an ordinary $4\frac{1}{2}$ V flat torch battery may be connected to the instrument with a plug.

To produce a simple circuit with a flat, reproducible characteristic (to which corrections, if required, can readily be applied), the successive amplifier stages are *RC* coupled (*fig. 3a*). No transformers are used, thus saving space, weight and costs. The moving-iron microphone is connected directly to the first amplifier stage, and the earphone (also of the moving-iron type) is directly connected to the output transistor. The volume is controlled by a potentiometer R_3 after the second amplifier stage. To prevent the threshold of pain from being exceeded, the maximum power of the output stage can be limited by a variable resistor R_1 in the collector circuit of that stage (this will be dealt with later). A multi-position switch, comprising S_1, S_2 and S_3 in fig. 3, serves for switching the battery on and off and for attenuating the low and high tones. The microphone can be switched over to a listening coil, which enables signals to be picked up inductively from the field of an exterior coil, such as that of a telephone receiver. The listening coil may also be used in combination with a loop circuit fitted in theatres and other buildings for the benefit of the hard of hearing [4]).

Details of the amplifier

The functioning of the circuit can best be understood by considering the D.C. and A.C. circuits separately. The D.C. circuit is shown in fig. 3*b*; it is obtained from fig. 3*a* by the omission of all branches containing capacitors. The A.C. circuit is represented in simplified form in fig. 3*c*, in which all capacitors acting as short-circuits to the alternating signal current are considered as ideal short-circuits and the internal resistance of the supply battery is neglected. Accordingly, resistor R_5 has also been omitted in fig. 3*b* and *c*. (R_5, together with C_5, provide decoupling, that is, they prevent alternating voltages, which appear across the internal resistance of the battery,

[4]) See the article quoted in note [3]), page 42.

from being fed back into the preceding amplifier stages.)

The transistors in the first three amplifier stages are wired in the same way, both for A.C. and D.C. The resistors R_b and R_c in each of these stages serve for the D.C. biassing. With $R_b = 2.2$ kΩ and $R_c =$ 39 kΩ, the battery voltage being 1.3 V, the operating point P shown in *fig. 4* is obtained. This point is chosen very low in the family of characteristics (small I_c), the object being to conserve the battery by a minimum consumption of current (the collector current is about 0.4 mA, the base current about 10 μA).

Since resistor R_b is not connected directly to the battery, but via the resistor R_c, there is not only a signal (A.C.) feedback but also a biassing (D.C.) feedback. The biassing feedback is the more important, since this makes the collector direct current (I_c) less sensitive to temperature changes, which in turn prevents the amplification from varying appreciably with temperature changes of a few degrees centigrade. We shall now consider this subject in more detail.

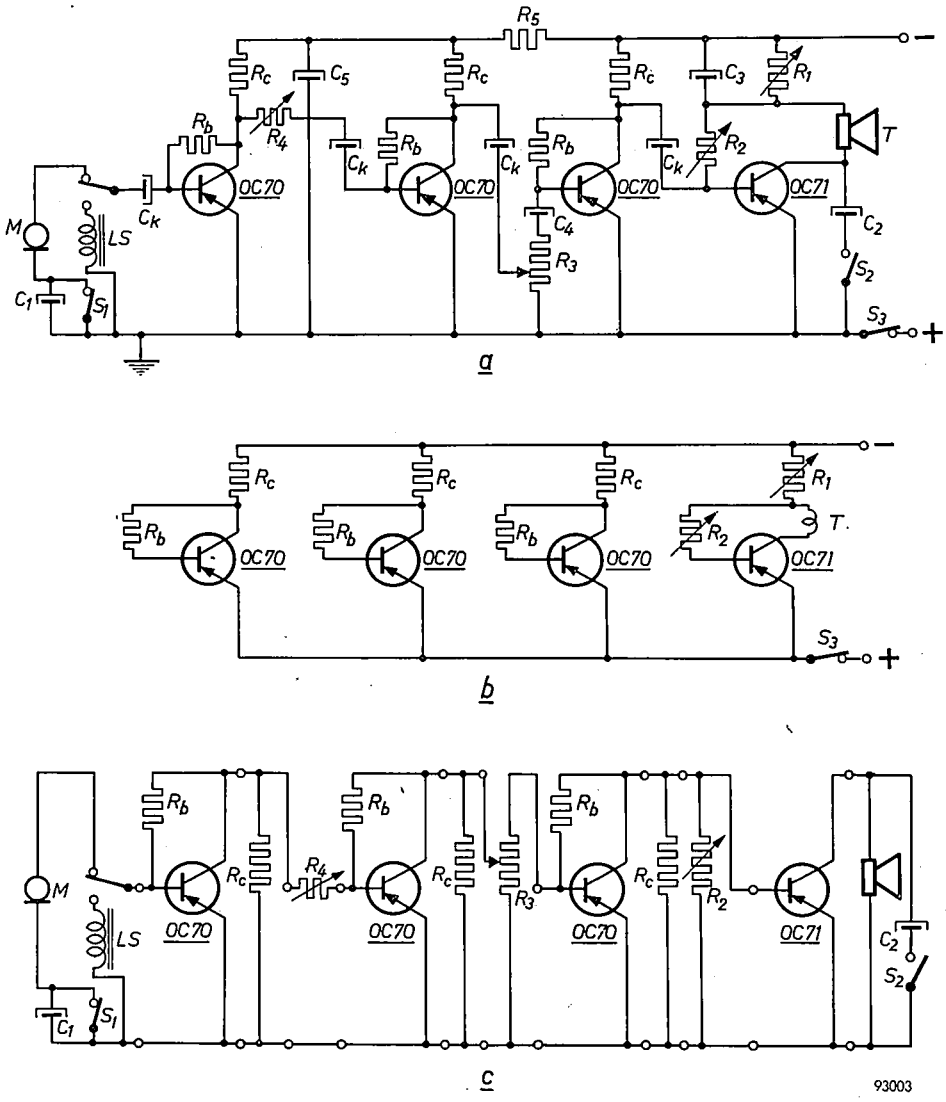


Fig. 3. *a*) Complete circuit diagram of hearing aid KL 5500. *b*) Direct-current circuit. *c*) Alternating-current circuit, drawn as a cascade arrangement of fourpoles, terminated at each end by a twopole.

M microphone; *LS* listening coil; S_1, S_2 and S_3 switches, for attenuating the low and high tones, and for switching the apparatus on and off. S_1, S_2 and S_3 are combined in a four-position switch. In the first position, S_3 only is open (apparatus switched off), in the second position all contacts are closed (high tones attenuated), in the third position S_2 is open (all tones amplified normally) and in the fourth position S_1 is also open (low tones attenuated).

Stabilizing the amplification against temperature variations

It can be seen in fig. 4 that the characteristics for constant base current I_b are almost horizontal. We may therefore write, to a good approximation [5]),

$$I_c = I'_{c0} + \alpha' I_b. \quad \ldots \ldots \quad (1)$$

The two terms of which I_c is composed are indicated in fig. 4; α', which is the current amplification factor, is a constant in so far as the lines for constant I_b are equidistant. What does not appear from fig. 4 is that I'_{c0} is strongly dependent on the temperature: for every 10 °C increase in temperature, I'_{c0} is approximately trebled [6]). The current amplification factor, on the other hand, is only slightly sensitive to temperature variations. If I_b is constant, changes in I'_{c0} will entail the same changes in I_c (see (1)). Since according to fig. 4, I'_{c0} and $\alpha' I_b$ in the first three amplifier stages are of the same order of magnitude, the appreciable relative changes of I'_{c0} with temperature variations of a few degrees will have a considerable effect on I_c. If R_b were connected directly (not via R_c) to the battery,

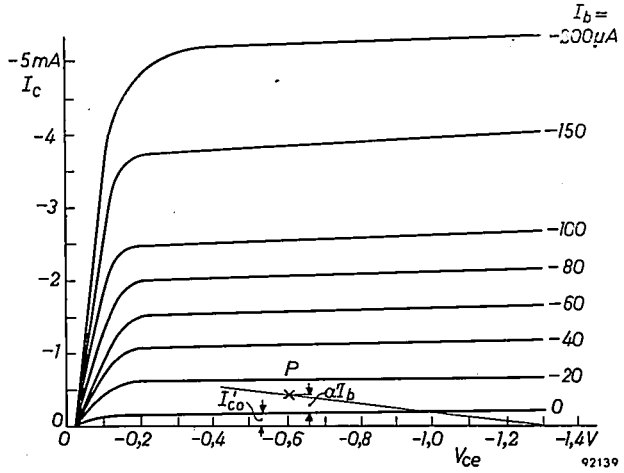


Fig. 4. Collector current I_c as a function of the voltage V_{ce} between collector and emitter, with the base current I_b as the running parameter, plotted for transistor OC 70, as used in the first three amplifier stages. The direct-current operating point is chosen at point P.

I_b would indeed be constant. This may be seen from *fig. 5a* which is the D.C. circuit for such an amplifier stage: $I_b = V_0/R_b$, which is constant (the voltage of approx. 0.1 V between base and emitter is negligible compared with the battery voltage V_0). However, in the circuit with negative feedback (fig. 3*b*), I_b depends on I_c. If I_c rises with rising temperature, owing to an increase in I'_{c0}, the voltage across R_b will decrease as a result of the increasing voltage drop across R_c. For this reason the current I_b, and

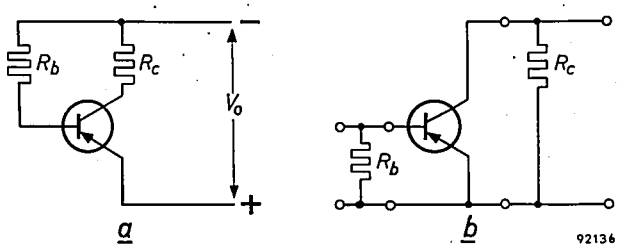


Fig. 5. *a*) The direct-current circuit of an amplifier stage, for the case that the resistors R_b of fig. 3*a* are connected directly to the battery; (*b*) alternating-current circuit.

hence $\alpha' I_b$, will also decrease, and therefore I_c will change less than I'_{c0} (see (1)). In this way a measure of stabilization of I_c is achieved.

For a D.C. circuit as in fig. 5*a*, the circuit for A.C. would be as shown in fig. 5*b*. It can be seen that now the transistor by itself occurs as a fourpole (shunted by resistors). Since the output of each transistor would then be virtually short-circuited by the input of the following transistor, the behaviour of the transistor as a fourpole is very simple, for we may take for the input resistance and the current amplification the values obtaining with short-circuited output (see pp. 23-25 and figs. 11 and 12 of the article quoted in [1])). This input resistance decreases sharply with increasing I_c [7]). If I_c rises with rising temperature, a larger portion of the output current of any stage will therefore flow into the base of the following transistor, and the output current of that transistor will increase in the same proportion. The total gain of the four stages would consequently drift by 7 to 8 dB per 10 °C, if it were not for the feedback actually applied, which reduces the temperature variations of I_c by a factor which in the present case is 2.5. If we wish to ascertain to what extent this will reduce the influence of temperature on the amplification, we must bear in mind that the stages in fig. 3 are rather more intricate than those in fig. 5, not only for direct current, but also for alternating current. In fig. 3*c* it can be seen that the transistors actually form fourpoles only when considered *together* with the resistors R_b. This makes the situation somewhat

[5]) The fact that the characteristics do show a certain slope, small as it may be for the scale values used here, makes itself perceptible only if R_c is much larger than the 2.2 kΩ used here. In that case a small change in I_c is associated with a large change in V_{ce}, and formula (1) can then no longer be used.

[6]) An article on temperature effects in transistors is shortly to be published in this Review. *Ed.*

[7]) See, for example, J. P. Beijersbergen, M. Beun and J. te Winkel, The junction transistor as a network element at low frequencies, II. Equivalent circuits and dependence of h parameters on operating point, Philips tech. Rev. **19**, 98-105, 1957/58 (No. 3), particularly page 104, in which the input resistance with output short-circuited is denoted h_{11}^e.

complicated. It will be enough here to report that the total gain of the hearing aid under discussion increases by only 2 to 3 dB per 10 °C rise in temperature.

Although the signal negative feedback reduces amplification per stage, it makes the amplification less dependent upon the properties of the individual transistors. Since individual transistors of the same type show a considerable spread in their properties. such stabilization is very welcome: the transistors in the first three stages can now be replaced without any re-adjustments being necessary.

Another adverse influence on the amplification is the fact that the coupling resistors R_c constitute parasitic loads on the transistors. The advantages of simple circuitry without transformers, and with stabilization against temperature variations, entail quite a considerable sacrifice of gain in the pre-amplifier. The theoretical maximum power gain of an OC 70 transistor — i.e. the power gain with ideal matching at the input and the output — is about 36 dB at the operating point considered. In the actual circuit, however, the first two amplifier stages produce only 19 dB each and the third 20 dB. The final stage produces 22 dB, making an overall (electrical) gain of about 80 dB.

The output stage

The small electromagnetic earphone is connected directly in the collector circuit of the output transistor OC 71. The variable resistor R_1 (fig. 3*a* and *b*), used for limiting the power output, is normally set at zero. It is moreover bypassed by a capacitor C_3, so that no signal feedback occurs and the output transistor forms, in itself, a fourpole (fig. 3*c*). The output characteristics of this transistor (with the input current I_b as the running parameter) are shown in *fig. 6*.

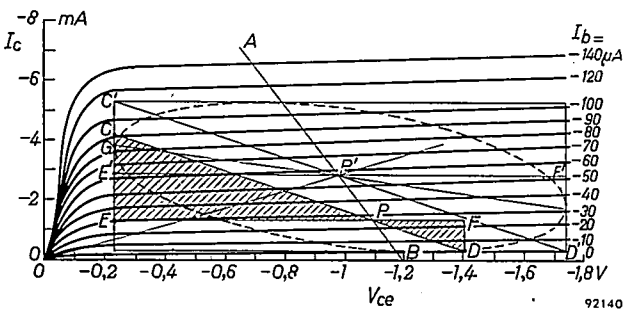


Fig. 6. Output characteristics of the output transistor OC 71. *B* indicates the battery voltage (taken as 1.2 V). The slope of *AB* corresponds to the D.C. resistance of the earphone (75 Ω). With a sinusoidal voltage across the load impedance (the earphone) an ellipse is drawn around the operating point; the diagonal of the coordinate rectangle tangential to this ellipse has a slope that corresponds to the value of the load impedance in ohms. To determine the most favourable operating point and load impedance, only this diagonal needs to be considered.

Biassing. The earphone has a D.C. resistance of 75 Ω. If $R_1 = 0$, the operating point then lies on the line *AB* (fig. 6) of slope corresponding to 75 Ω, running through the point *B* which represents the battery voltage. By varying the base direct current with R_2 (fig. 3*a* and *b*), the operating point can be shifted along *AB*. The position of the operating point is chosen with a view to conserving the battery, that is to say it is chosen as low as is compatible with the power required to be available at the earphone. This power, as already stated, is 1 mW for normal purposes. This is the power we wish to have available when using a battery consisting of one cell, that is, at a battery voltage of about 1.2 V (to which the voltage drops when the battery is run down). The family of characteristics is limited on the left (the curves drop sharply) and underneath (since the collector current cannot change sign). Imagining, for the sake of simplicity, the earphone impedance to be replaced by a pure resistance, and assuming that *P* is the operating point, we can draw through *P* a load line *CD*. The area of the triangle *PDF*, which is the smallest of the two hatched triangles *PDF* and *PCE*, is then a measure of the available useful power, i.e. the power that the transistor can deliver without serious distortion. If the operating point is moved up along *AB* by raising the base current, the areas of the two triangles will then approach each other; they become equal when CD has moved to *C'D'*. If the slope of *CD*, i.e. the value of the load, is chosen such that the available useful power is exactly 1 mW in the situation corresponding to *C'D'*, we have then found the most economical operating point at the given voltage. In our case the required load is found to be about 300 Ω. The slope of *CD* in fig. 6 has been chosen in accordance with this value.

The situation in reality is complicated by the fact that the earphone constitutes a strongly inductive load. For a sinusoidal output signal of a single frequency, the output characteristic is represented by an ellipse drawn around the operating point, instead of a straight line. The diagonal, from top left to bottom right, of the coordinate rectangle tangential to this ellipse, has a slope that corresponds to the impedance of the load. If this diagonal does not extend to regions where distortion may be expected, then neither will the ellipse extend to such regions. The value of 300 Ω mentioned above is therefore the value which the impedance of the earphone (and not the purely resistive portion of this impedance) must have in order that the biassing be such as to give the most economic operating point. Of the available power only that part given

by the area of triangle $P'E'G$ is dissipated in the earphone.

Since the impedance of the earphone depends on the frequency, it can only be exactly 300 Ω for one frequency. Earphones are used which have this impedance at a frequency of 1000 c/s.

A second complication is that no definitive conclusion can be drawn from fig. 6 regarding the distortion to be expected. It is not at all evident, for instance, that serious distortion occurs if the transistor is driven to very small values of I_c. This distortion is due to the fact, already mentioned, that the input resistance of the transistor increases sharply at small values of I_c. Nevertheless, the method indicated above of determining the most economical operating point and the most favourable load impedance from fig. 6 does lead to results of practical value.

Since the current amplification factor of the OC 71 transistor may vary between 30 and 75 with individual transistors of the same type, the base current at which the most favourable operating point is obtained depends upon the transistor employed. It must therefore be possible to adjust the base current during assembly and subsequently if replacement of the output transistor should be necessary; this is the reason for making the resistor R_2 variable.

Limiting. The operating point discussed above automatically provides that, with increasing input signal, the power increases only very little from the moment that serious distortion sets in. This is due to the fact that the limiting of the output signal and the serious distortion associated with it both occur equally on both sides of the operating point (symmetrical limiting). In this way the output power is prevented from exceeding a certain upper limit, thereby providing an effective safeguard against "acoustical trauma".

The level at which the limiting becomes operative can be regulated with resistor R_1 (fig. 3*b*). The higher this resistance, the lower the collector voltage. This voltage is also across the base resistor R_2 (the slight potential difference between base and emitter is negligible) and thus the base current falls proportionately when R_1 is increased. In its turn, the collector current I_c decreases proportionately with the base current, at least as long as $I_c \gg I'_{c0}$ (see (1)). The operating point is thus shifted along the line OP' (fig. 6) towards the origin; the symmetrical cut-off is thereby maintained.

Since the amplification of the output stage is not stabilized by negative feedback, the considerable spread in the current amplification factor of individual output transistors would normally appear in full in the total gain. To prevent this happening, a variable resistor R_4 is introduced between the first and second stages (fig. 3*a* and *c*) with which the total gain can be adjusted during assembly or after the replacement of the output transistor. This also compensates for the residual spread in the characteristics of the earlier stages, still remaining in spite of negative feedback.

As regards direct current, too, there is no feedback in the output transistor when R_1 is set at zero, and therefore no stabilization of the operating point against temperature variations. This is not necessary, however, since the collector bias current I_c of the output transistor is adjusted to a much higher value than that of the transistors in the first three stages. The contribution of I'_{c0} to I_c is therefore relatively much smaller (see (1)).

If the battery voltage is doubled or trebled by using a battery consisting of two or three cells, the operating point moves along OP' away from the origin. The maximum amplitudes of current and voltage are also doubled or trebled and the available useful power is increased by a factor of 4 or 9, as the case may be. Thus, with a battery of three cells the available power is brought up to the 10 mW that is necessary for some cases of severe deafness and when bone conduction is employed. Raising the battery voltage causes an increase not only in the biassing current of the output transistor but also in that of the transistors in the first three stages. Since higher operating currents entail lower input resistance, the loss in the resistors R_c (fig. 3*c*) is lower. This causes the total gain acoustic to increase from 55 dB for one battery cell to 63 dB and 67 dB for two and three cells respectively.

With all three battery voltages the resistor R_1 allows the ear specialist to lower the maximum output power (the ceiling) continuously by an amount from 0 to 20 dB. It can be seen in fig. 3*b* that the limiting of the output is associated with a reduction in battery current, as one would wish.

Microphone, earphone and response curves

The input resistance of the first amplifier stage is relatively low, being somewhere between 1000 Ω and 1800 Ω, and thus readily permits the direct connection of a moving-iron microphone. Crystal microphones, as normally used in valve hearing aids, call for load impedances of the order of 0.1 to 0.5 MΩ. This would necessitate the use of a matching transformer, and even then less power would be delivered, with more noise. The moving-iron microphone used in the KL 5500 hearing aid has an internal resistance of 1000 Ω and is particularly suitable for

use with transistor hearing aids. Its smooth frequency characteristic facilitates the attainment of the desired overall response curve. The microphone functions as follows [8]). The vibrations of an aluminium diaphragm are transferred to an armature composed of a material of high permeability. This armature constitutes the "galvanometer" diagonal of a magnetic "Wheatstone bridge". The resistance arms are formed by air-gaps in the magnetic system, and the requisite magnetic flux is supplied by a small permanent magnet forming the other diagonal of the bridge. The magnetic flux through the armature is dependent in direction and magnitude upon its deviation from its equilibrium position. A coil fitted around the armature converts the flux variations into an alternating voltage. The system is extremely sensitive: at 1 kc/s the sensitivity is about 0.3 mV/μbar when the microphone is terminated by 1000 Ω. By means of suitably dimensioned resonant air cavities the frequency characteristic can be made fairly flat between 400 and 3000 c/s, which is the important range for speech (*fig. 7*). The consequence of these measures is that the characteristic falls more sharply outside this range, which has the advantage,

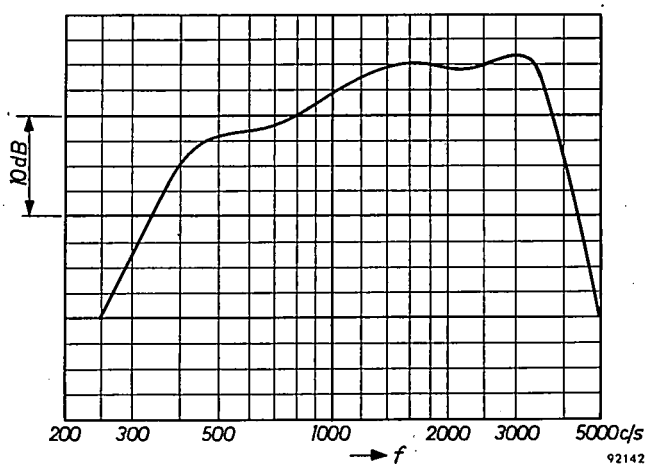


Fig. 7. Frequency characteristic of the microphone. The quantity actually plotted is the voltage across a resistor of 1000 Ω terminating the microphone, the latter being in a sound field of constant pressure.

where the lower frequencies are concerned, of very efficiently suppressing intermodulation phenomena, which may occur in the presence of "boom", in motor vehicles, for example.

The microphone casing is surrounded by two rings of a metal with a high initial permeability, the purpose of which is to shield the microphone against stray magnetic fields (attenuation by about 16 dB) and to help create a quiet background.

As regards the earphone, the patient has a choice of three types. The type Ph 1 earphone has an almost flat frequency characteristic (*fig. 8*). Unavoidable

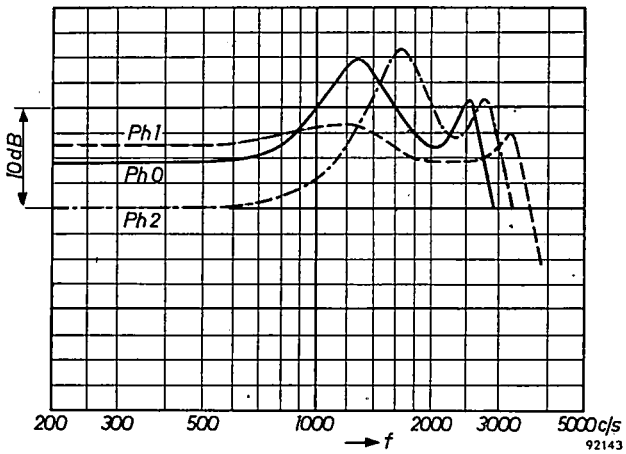


Fig. 8. Frequency characteristics of the three earphones (Ph 0, Ph 1, Ph 2) from which patients may choose. As a function of the frequency f of the constant current through the earphone, the sound pressure is plotted as measured in an artificial ear of 2 cm³ volume in which the earphone was fixed.

resonances have been attenuated as much as possible by the introduction of damping. Fig. 8 also shows the characteristics of the two other earphones, types Ph 0 and Ph 2. In these types the resonances are attenuated to a lesser extent.

With the sum of the frequency characteristics of microphone, amplifier and earphone, an overall response curve (fidelity characteristic) is obtained as shown in *fig. 9*, using an earphone of the type Ph 1. By switching C_1 or C_2 (fig. 3), a portion of the low or

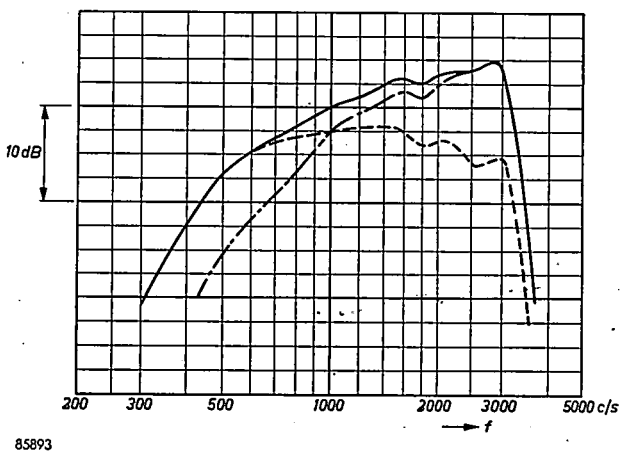


Fig. 9. Overall response curves (fidelity characteristics) of the complete hearing aid KL 5500 with earphone Ph 1, for the three positions of the control switch. The quantity plotted is the sound pressure measured in an artificial ear of 2 cm³ volume in which the earphone was fixed, the microphone being in a field of constant pressure.

Full curve: S_1 closed, S_2 open (see fig. 3c); dashed curve: S_1 and S_2 both closed; dot-dashed curve: S_1 open, S_2 open.

[8]) For a detailed description see: B. B. Bauer, A miniature microphone for transistorized amplifiers, J. Acoust. Soc. Amer. **25**, 867-869, 1953.

high frequency range, respectively, can be cut off, producing the three curves illustrated. These curves can also be changed by the choice of earphone, so that many combinations are possible.

Case noise

The extremely low battery costs make it possible to keep this hearing aid in continuous use. This being so, it is more important than ever that the user should be able to hear with the least possible effort. Various factors are involved here, an important one being, as we have seen, that the tone quality perceived by the deaf person should be as agreeable as possible, heard against a quiet background. The most troublesome kind of background interference heard by users of hearing aids is case noise. This is caused by friction between clothing and the case of the instrument, or by friction between a piece of clothing stretched over the case and an adjacent fabric rubbing over it. Modern hearing aids are so smoothly finished, and there is so little movement of clothing with respect to the case, that the first source of case noise may, for all practical purposes, be neglected. The movement between adjacent fabrics bearing on the case, however, is many times greater in amplitude and is always present. It has the nature of "white noise"; that is to say, all frequencies in the band passed by the hearing aid are represented in almost the same intensity. Sufferers from certain types of deafness may find this extremely troublesome, and it is therefore very important to suppress this kind of interference as effectively as possible.

The vibrations produced by case noise can reach the microphone in three ways.

a) The friction may set the case and the chassis of the instrument in vibration; the vibration is transmitted to the microphone via the microphone mounting. This transmission is almost entirely determined by the natural resonance of the system consisting of the microphone (mass) and its elastic mounting (stiffness). The natural frequency can easily be kept below 100 c/s, in which case the amplitude of vibrations within the speech range remains negligible.

b) The vibrations of the case can be transmitted to air cavities inside the hearing aid. These cavities often resonate and thereby pass certain frequency ranges with extra intensity. The vibrations may possibly reach the microphone via these cavities.

c) The vibrations of the clothing material reach the microphone diaphragm via the air in the normal way.

An apparatus has been developed by Philips for making comparative measurements of the sensitivity of different hearing aids to case noise. The results of such measurements make it possible to assess objectively the measures taken to suppress interference due to case noise.

The apparatus contains an endless cotton belt driven by a small motor. The hearing aid to be tested is pressed against the moving belt, either with or without an intermediate piece of clothing fabric stretched over the case. The case noises thus produced are reproducible, and in a certain sense "standardized". The earphone of the hearing aid is fixed in an "artificial ear", and the voltage produced in the microphone of this device is fed to a voltmeter via a band-pass filter (bandwidth one third). The case noise is subsequently stopped and a standardized "acoustic" noise substituted, i.e. a noise whose vibrations reach the instrument's microphone only by the normal acoustical means, that is via the air and not by the means mentioned under *a*) and *b*) above. This acoustic noise is produced by a loudspeaker set up nearby the hearing aid and fed by a noise generator. The volume of the acoustic noise is adjusted such that the deflection of the voltmeter is the same as when the case noise was operative. A measurement is then made of the sound pressure of the acoustic noise field at the position of the hearing aid; this is done with the aid of a microphone connected via the same filter to a voltmeter which is calibrated in decibels above a threshold value of 10^{-4} dyne/cm^2. In this way the sound pressure is ascertained of the acoustic noise which, within a frequency band of one third, produces just as much interference as the standardized case noise. By repeating the measurement with different band-pass filters of one third, the entire

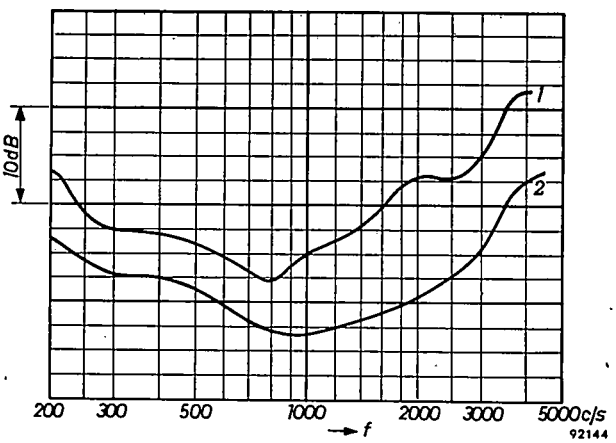


Fig. 10. Case-noise interference as a function of frequency, *1* with microphone opening partly covered, *2* with opening uncovered.

audio frequency range can be covered. The advantage of this method is that the results depend only upon the extent to which the case-noise vibrations reach the microphone of the hearing aid, and not upon the amplification or the response curve of the instrument.

The method can also be used to investigate the effect of the conditions under which the case noise is generated. As an example, *fig. 10* shows the effect of partly covering the microphone opening by vibrating fabric. With a given amplitude of vibration, the variations in air pressure in the small cavity in front of the microphone diaphragm are larger the more completely the opening is covered. It is seen from fig. 10 that when the opening is partly covered the interference is greater at all frequencies than when the opening is uncovered; in the range of about 1000 to 3000 c/s the difference is as much as 10 dB. It is important, therefore, when wearing a hearing aid, to keep the microphone opening entirely free.

The results of case-noise measurements have led to the adoption of various measures in the mechanical construction of the type KL 5500 hearing aid — and in other types too — to suppress this interference. It has been found that apparently minor modifications often have a substantial effect.

Summary. The use of transistors in the KL 5500 hearing aid has made it possible to reduce battery costs drastically and at the same time to increase the available output power and amplification. As a result, the apparatus can serve a wide circle of deaf persons. After a discussion of the various requirements to be satisfied by a hearing aid, a description is given of the amplifier, in which four transistors are used. The effects of temperature variations and of the spread in the properties of individual transistors are largely eliminated by means of negative feedback. The highly sensitive moving-iron microphone is directly coupled to the input stage, without the intermediary of an input transformer. A transformer is also superfluous for the earphone, since with transistors the maximum available power is obtained with a load having an impedance of a few hundred ohms, which is the normal impedance for electromagnetic earphones. The variable limiter prevents the maximum output power from exceeding the threshold of pain. The article concludes with some details of investigations into the interference caused by case noise.

AN APPLICATION OF TELEVISION FOR THE DISCOVERY OF VARIABLE STARS

621.397.611.2:523.841.3

Variable stars, i.e. stars which vary in brightness, are of great interest in astronomy, one reason being that it is often possible to determine their distance from the earth. An electronic apparatus to help in the discovery of variable stars has been constructed at the Kapteyn Astronomical Laboratory at the University of Groningen. The apparatus makes use of techniques used in television [1]).

Variable stars may be detected by means of two photographs of the same region of the sky. The photographs are taken at different times but, apart from this, the circumstances should be as similar as possible. If the brightness [2]) of a star has changed in the lapse of time between the two exposures, the change is primarily to be detected in a difference of diameter between the photographic images of the star.

Comparison of the photographs is generally carried out by means of a "blink-microscope". With this instrument the same region of the two photographs is seen in turn in the field of vision. No change is then seen in stars having constant brightness, but variable stars appear to grow and shrink, as it were. With this procedure, however, the number of stars missed is fairly large, since it is often necessary to re-examine the plates several times before any variation in the image is noticed and not always the same stars are retrieved when a certain pair of plates be examined more than once.

Various other techniques have been devised for the search for new variable stars. One is a comparative method involving the original negative of one exposure and a positive of the other; the two are examined in conjunction, the black spots on the former being made to cover up the "holes" in the latter. The result is that stars of constant brightness become invisible, while variable stars become visible as a result of the slight differences in their two images. However, the photographic preparations that this method requires present considerable difficulty.

The new apparatus is based on the same idea, but the photographic preparations are eliminated. With the aid of the flying-spot scanner as used in television, it is possible to display a photographic negative on a CRT screen either as a positive or as a negative image, whichever is desired. The flying-spot scanner embodies a special kind of cathode-ray tube having a small bright spot with hardly any afterglow. The spot traces a raster or pattern of lines on the (flat) screen in exactly the same way as in a television picture tube. A lens projects the image of the spot onto the transparent object that is being examined; in other words, the spot is made to scan the object. The light that passes through the object is conveyed by a condenser lens to a photo-multiplier tube which delivers a signal proportional to the light passed by the object at that particular point of its surface. The signal can be used to modulate the intensity of the electron beam in a picture tube, the screen of which will accordingly display an image of the object [3]).

With the aid of a half-silvered mirror it is possible to scan two objects at the same time using one cathode-ray tube. Such an arrangement is used in the present apparatus.

Fig. 1 shows the principle of the apparatus. A half-silvered mirror M_4 and three highly polished mirrors M_1, M_2 and M_3 split the light from the type MC 13-16 cathode-ray tube K into two beams; lenses L_1 and L_2 form images of the spot on both F_1 and F_2, which are photographic plates of the same region of the night sky, of the kind referred to above. The signals from the two multiplier tubes P_1 and P_2 are electronically subtracted and the difference signal, after amplification, is employed as a picture signal for modulating picture tube B. In fact, the sign of one signal is reversed, but not that of the other; the two are then added. If each of the signals is taken separately, therefore, one will produce a negative image on the screen, and the other a positive image, as shown in *fig. 2*.

If two identically similar plates are examined in this way (care naturally being taken that the plates are so positioned that corresponding points are always scanned at the same time), the signals from the two photo-electric multiplier tubes will always be equal. Consequently the difference signal will remain zero and nothing will be seen on the screen.

---

[1]) J. Borgman, Dissertation, Groningen 1956.

[2]) In astronomy the term "brightness" is taken to indicate the illumination intensity produced on earth in a plane perpendicular to the line of sight to the star.

[3]) For a more detailed discussion of the flying-spot scanner, see F. H. J. van der Poel and J. J. P. Valeton, Philips tech. Rev. **15**, 221-232, 1953/54. For details of the cathode-ray tube, see A. Bril, J. de Gier and H. A. Klasens, Philips tech. Rev. **15**, 233-237, 1953/54.

Should the two plates differ in some way (e.g. if they carry the images of a star whose brightness has changed in the time between the two exposures), the picture signal will not be zero and the difference between the plates will be visible on the screen as a patch that is either lighter or darker than its surroundings. An example appears in fig. 2.

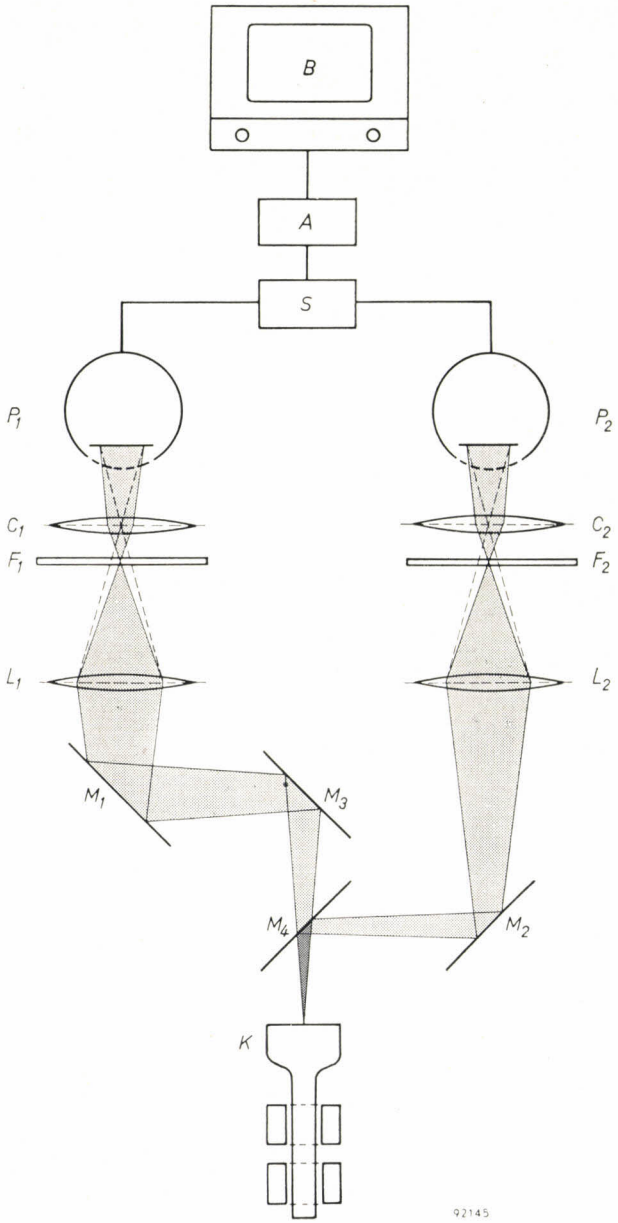


Fig. 1. The principle of the two-channel flying-spot scanner. By means of the half-silvered mirror M_4 and the highly polished mirrors M_1, M_2 and M_3, the light from the cathode-ray tube K is split into two beams. Images of the raster traced on the screen of K, reduced about ten times, are projected by L_1 and L_2 onto the two photographic plates (F_1 and F_2) to be compared. C_1 and C_2 are condenser lenses. The signals delivered by photo-electric multipliers P_1 and P_2 are subtracted. After amplification, the difference signal is made to modulate the intensity of the electronic beam of picture tube B, which forms part of a normal television set. The sweep voltages for the picture tube and for the flying-spot cathode-ray tube are synchronized. Details of the circuits may be found in the author's dissertation (see footnote [1]).

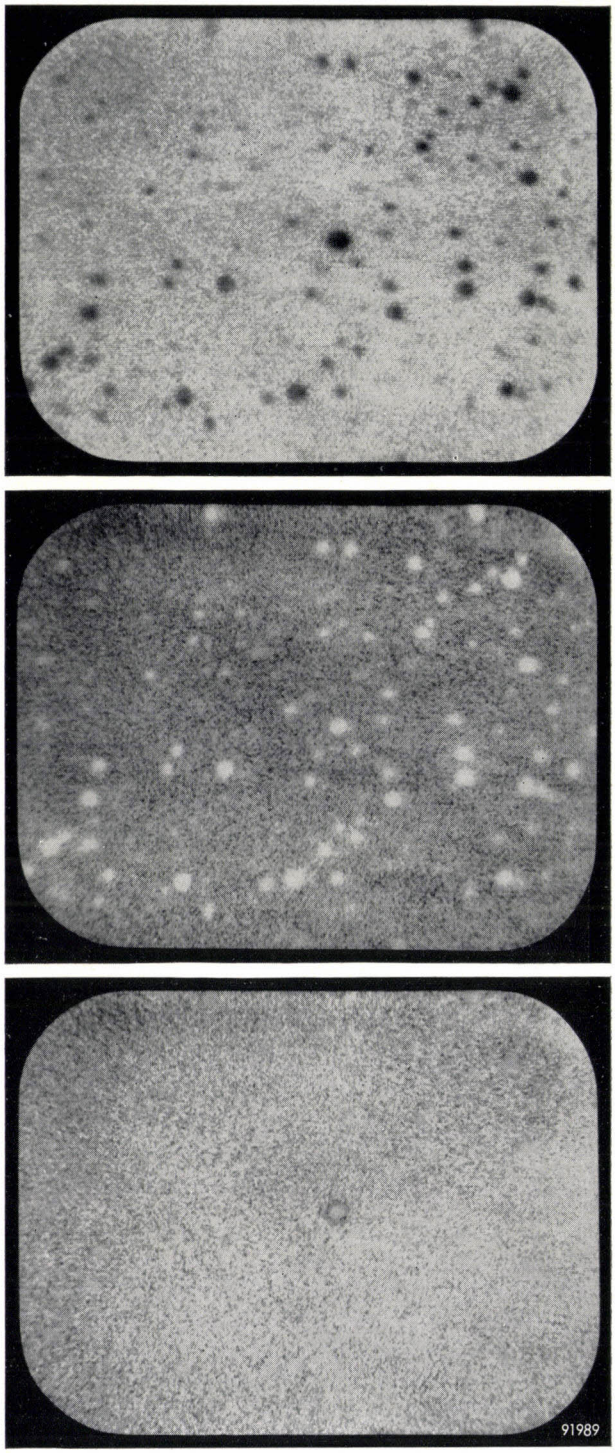


Fig. 2. Photos of the television screen taken during the scanning of a region of the plate containing a variable star. *Above:* "Negative" image of one plate. *Centre:* Reversed or "positive" image of the other plate. *Below:* The sum of the two images above. A clear case was selected in order that an unretouched photo might be shown. (Photos by Central Photographic Service of Groningen University.)

Fig. 3 is a photograph of the complete apparatus.

The chance of finding a variable star on a given pair of plates with the aid of this instrument cannot easily be expressed as a number; it depends both on the *brightness* of the star and upon the *change*

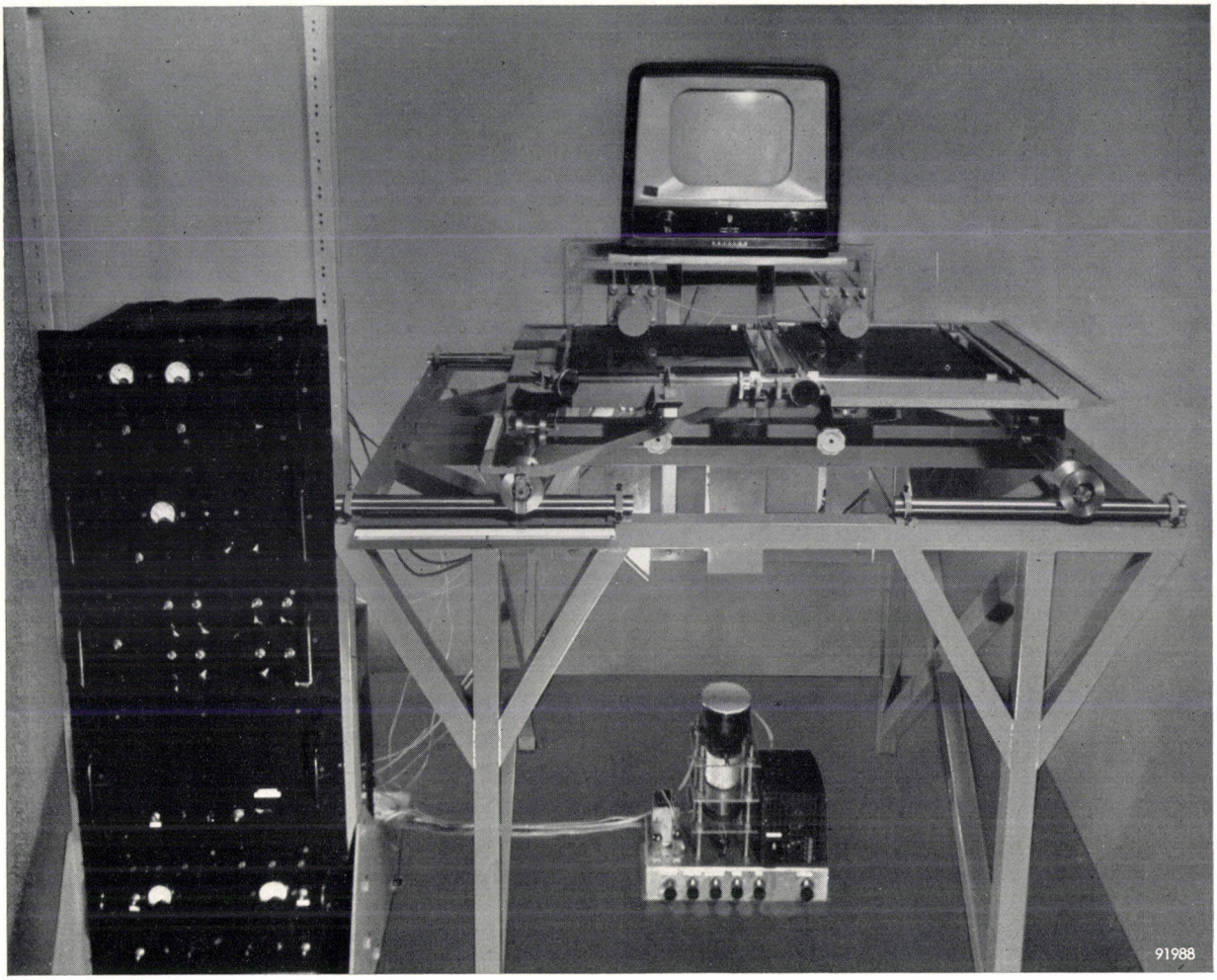


Fig. 3. The apparatus described in the text. The plate holder can be moved on rails in two directions at right-angles to each other. The two multiplier cells can be seen under the television set. The chassis carrying the cathode-ray tube and sweep generators is on the floor. The remaining electronic equipment is accommodated in the rack. (Photo by Central Photographic Service of Groningen University.)

in brightness (i.e. the difference in the photographic images). Suffice it to say that the chance is greater than when a blink-microscope is used, and that there is also considerable saving of time.

An instrument of this kind has general utility for the detection of slight differences between transparent objects. A further case in astronomy is for the detection of heavenly bodies whose positions relative to the "fixed" stars change (asteroids in particular). Outside astronomy, the instrument might be used for detecting forgeries (of bank notes, for example) and differences between air photos of the earth's surface made at different times.

J. BORGMAN *).

*) Kapteyn Astronomical Laboratory, University of Groningen.

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.

2464: B. B. van Iperen: Klystrons reflex pour ondes millimétriques (Le Vide **11**, 264-266, 1956; No. 65). (Reflex klystrons for millimetre waves; in French.)

Three reflex klystrons for wavelengths of 12, 8 and 4 mm are described. These valves, which are essentially of the same construction, have a mechanical tuning range of 10-25% and deliver a continuous power output of the order of 100 mW. The measured values of the output power are found to be nearly 2/3 of those expected from the theory for all three types of valves. Using this correction factor, which is assumed to be due to the influence of space charge and transit time spread, an estimate is given of the limiting frequency for valves of the type described. It is found that the maximum frequency of 82 000 Mc/s obtained with the 4 mm valve is about one half of this limit.

2464a: F. Coeterier: Tubes à réflexions multiples (Onde électr. **36**, 917-919, Nov. 1956). (Multireflection tubes; in French.)

Some new types of multireflex klystrons are described. The basic problem of forming electron bunches and making them oscillate with constant phase around an interaction gap is solved in two ways: firstly the known method of a special retarding field, and secondly a combination of drift-tube action with a retarding field, using three interspaced gaps. Tunability is achieved by a moving metal bar between the Lecher strips forming the resonant system. If a quick sweep of the available bandwidth is required as e.g. in marine radar beacons, this metal bar can be replaced by a vibrating vane. The output is of the order of 10 W with an efficiency of 20% for both types.

2464b: J. G. van Wijngaarden: Possibilities with disc-seal triodes (Onde électr. **36**, 888-892, Nov. 1956).

With the introduction of the L-cathode in disc-seal triodes, tubes for higher frequencies than possible with normal oxide-coated cathode can be constructed. Also the H.F. output power can be considerably greater. The frequency limitations are imposed by the circuitry around the tube (especially the possibility of building a quarter-wavelength cavity around the grid-anode region) and by the transit time of the electrons. The latter is, for a given cathode-grid distance, mainly a function of the admissible current density. The H.F. output power is a function of the current density and of the admissible plate and grid dissipations. Some results obtained with a triode for 4000 Mc/s and output power of more than 20 watts at 100 Mc/s bandwidth are discussed. The construction of triodes for frequencies as high as 10 000 Mc/s, with reasonable gain-bandwidth and power-bandwidth qualities, seems to be feasible.

2464c: S. Woldring: Over de ademhaling tijdens het spreken van dove kinderen (T. voor Doofstommenonderwijs **26**, 153-160, 1956, No. 4). (On the breathing of deaf children while speaking; in Dutch.)

In speech a good coordination between breathing and movement of the speech organs is important. In persons of normal hearing this coordination is achieved by listening. Investigation by means of so-termed pneumograms, taken on the chest and abdomen of normal and deaf children during rest and during speech, clearly demonstrate the lack of control of respiration in deaf children during speech. An improvement in the perception of sounds automatically brings about an improvement in speech.

2464d: H. Bremmer: Remarks on the connection between mode theory and ray theory (Nuovo Cimento (10) **4**, suppl. No. 4, 1552-1558, 1956).

By calculating a simple model for the propagation of radiowaves via the ionosphere the connection is shown between the description using the ray concept and the description with eigen functions. With the aid of the ray theory an integral formula is derived which contains the two reflexion coefficients (on earth and on ionosphere). This integral formula can be evaluated with the aid of the theorem of residues, which procedure yields the solution in terms of eigen functions. The connection between the two descriptions is due to the fact that the residues mentioned are determined by a resonance condition involving the two reflection coefficients, where the reflection coefficient on the ionosphere depends on the eigen functions of the problem.

2465: H. G. Bruijning: High-power pulse generators (T. Ned. Radiogenootschap **22**, 1-14, 1957, No. 1).

In the Philips Research Laboratory an experimental radar has been built, designed for pulses of 0.01 μs, at a repetition rate of 2000 pps. The generator had to deliver pulses of 15 kV, 15 A to the magnetron. This paper deals with the development of the pulse generator (200 kW pulses of 0.01 μs duration) and with the construction of an oscilloscope (bandwidth 400 Mc/s) specially built to be used in connection with the experimental radar mentioned above.

2466: S. Duinker: On the resolving power in the process of magnetic recording (T. Ned. Radiogenootschap **22**, 29-48, 1957, No. 1).

The recording process taking place in a thin magnetic layer situated at a certain distance in front of the recording head is analyzed if the head is magnetized by pulses superimposed upon a D.C. level, both with and without an additional A.C. biassing field. For both cases expressions are derived for the resolving power, defined as the limiting wavelength corresponding to the repetition frequency of extremely short pulses at which just once in a period the remanent magnetization inherent in the D.C. level is recorded on the tape travelling at a fixed speed. The resolution is shown to depend upon the ratio of pulse height to D.C. level, the shape of the field curve and the depth into the tape in the D.C. method, while in the A.C. method, the relative strength of the A.C. biassing field and the critical field strength of the tape are of importance also. It is demonstrated for the D.C. method that resolutions of the order of the gap length are obtainable while in the A.C. method a much smaller wavelength can be recorded.

2467: J. Bloem and F. A. Kröger: De diffusie van Cu in PbS (Chem. Weekblad **53**, 1-4, 1957, No. 1). (The diffusion of Cu in PbS; in Dutch.)

Experiments show that Cu has a strong preference for the Pb^{2+} lattice sites in PbS. At high temperatures ($T > 500$ °C) diffusion takes place mainly by a vacancy mechanism. For $T < 500$ °C, however, the lattice must be considered rigid; the diffusion of Cu occurs via the inter-lattice. Dependent on the atmosphere used (H_2 or H_2S) it is found that Cu can be made to diffuse into or out of PbS crystals at low temperatures.

2468: H. J. Oskam: Enkele aspecten van de hoogfrequente gasontlading (Ned. T. Natuurk. **23**, 1-15, 1957, No. 1). (Some aspects of the high-frequency gas discharge; in Dutch.)

The mechanism of the high-frequency gas discharge differs in a number of ways from that of the D.C. gas discharge; the former is in fact less complicated so that it is possible to set up a theory of the high-frequency discharge that agrees well with experiment. Energy transfer to the electrons is different in the two types of discharge and may be compared by the use of the so-called equivalent field. The γ-mechanism at the electrodes plays an essential role in the D.C. discharge, while in the high-frequency discharge its influence is not important. Whereas in the D.C. discharge the electrons disappear primarily as a result of the field, in the high-frequency discharge they disappear by diffusion to the walls, recombination in the gas and attachment. This is illustrated by comparing measurements of the breakdown fields in neon-argon for the D.C. case with those for the high-frequency case. The differences are explained in terms of the above remarks. By the application of high-frequency techniques it is possible to measure a number of gas-discharge quantities in a new way.

2469: J. A. Haringx: Design of corrugated diaphragms (Trans. Amer. Soc. Mech. Engrs. **79**, 55-64, 1957, No. 1.)

Three previous papers by the author set forth methods of calculating the rigidity of corrugated diaphragms, the stresses in the sheet material, and the non-linearity of the relation between load and deflection. As a further step, the introduction of a few simplifying restrictions having no fundamental effect on the problem leads to the concept of a chart giving at once the dimensions a diaphragm must have so as to conform to specific requirements. An example is included by way of illustration.

VOLUME 19, 1957/58, No. 5 pp. 145-176 Published 30th November 1957

Philips Technical Review

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

A 4000 Mc/s WIDE-BAND AMPLIFIER USING A DISC-SEAL TRIODE

by J. P. M. GIELES.

621.375.2.029.6:621.385.3.029.6

Although in recent years several special types of electronic tubes, such as travelling-wave tubes and klystrons, have been developed for use as amplifiers at very high frequencies, the oldest type of amplifying tube, the triode, can still be put to good use for this purpose. Following upon an earlier article, which described the design of a disc-seal triode, type EC 57, capable of operating at frequencies up to 4000 Mc/s, the present article describes an amplifier in which this tube can be employed. The amplifier is now being used, to give one example, for the microwave telecommunication network at present in course of construction in the Netherlands.

Introduction

In recent years work has been going on in many parts of the world on the construction of microwave radio links, intended for relaying multichannel telephony and television signals. A frequency band from 3800 to 4200 Mc/s (wavelength approx. 7½ cm) has been reserved for these links by international agreement. These frequencies, while being high enough to allow a satisfactory directional effect to be obtained without the use of too large reflectors, are not so high as to give rise to excessive fading in unfavorable weather conditions. In the present article a description will be given of an amplifier specially developed for the frequency range in question. The amplifier, which is at present being employed in the microwave telecommunication network in course of construction in the Netherlands, can be operated with two types of disc-seal triode, viz. type EC 56 and EC 57. With the EC 57 tube, which has already been discussed in this Review [1]), the amplifier can deliver a power of about 1.5 W. Although this is lower than the output at the same frequency of the multi-reflex klystron, also described on an earlier occasion [2]), the amplifier dealt with here will nevertheless be an important asset in fields of application for which the multi-reflex klystron is unsuitable.

The multi-reflex klystron can only be employed as an oscillator, for which reason it is used, for example, in beam transmitters with double frequency modulation [3]). Owing to the relatively low frequency of the auxiliary carrier wave, however, this system cannot be used when it is necessary to transmit signals with a very wide frequency band, as in television. In that case, only single modulation is possible, and when, as is usual in relay and terminal stations, the amplification is obtained at an intermediate frequency much lower than that of the signals to be transmitted, a mixer stage must be used to achieve the latter frequency. At centimetric waves a crystal diode is suitable for this purpose. Its output, however, being only a few milliwatts, is inadequate for establishing a reliable radio link. An amplifier is then needed which operates at the transmitter frequency.

Microwave amplifiers

Before describing the design of an amplifier for centimetric waves, we shall deal briefly with the main differences between amplifiers of this kind and those which operate at lower frequencies.

At frequencies above 500 Mc/s it is almost impossible to build resonant circuits composed of coils and capacitors. It is then necessary to resort to Lecher systems and, at frequencies above about

[1]) G. Diemer, K. Rodenhuis and J. G. van Wijngaarden, The EC 57, a disc-seal microwave triode with L cathode, Philips tech. Rev. **18**, 317-324, 1956/57 (No. 11).

[2]) F. Coeterier, The multireflection tube, a new oscillator for very short waves, Philips tech. Rev. **8**, 257-266, 1946.
F. Coeterier, The multi-reflex klystron as a transmitting valve in beam transmitters, Philips tech. Rev. **17**, 328-333, 1955/56.

[3]) See C. Ducot, Beam transmitters with double frequency modulation, Philips tech. Rev. **17**, 317-327, 1955/56.

2000 Mc/s, to resonant cavities. In the latter case the various stages of an amplifier are usually coupled by waveguides, and this fact must be taken into account in the design of the amplifier.

When using a triode in the centimetric wave range the only possible circuit is the grounded grid arrangement. The fairly low input impedance of the tube in this circuit entails a rather high driving power; this power has to be fed to the tube input via a waveguide. It is desirable that the tube should then be, as far as possible, a matched load to the waveguide. With centimetric waves it is hardly possible to use anything but a resonant cavity as the anode circuit, and this must be coupled to the output waveguide. The circuit must be designed such that the tube can provide the desired gain at the operating frequency, with a sufficiently large bandwidth. Moreover, the anode resonant cavity must be tunable. We shall deal in this article with various problems arising in this connection.

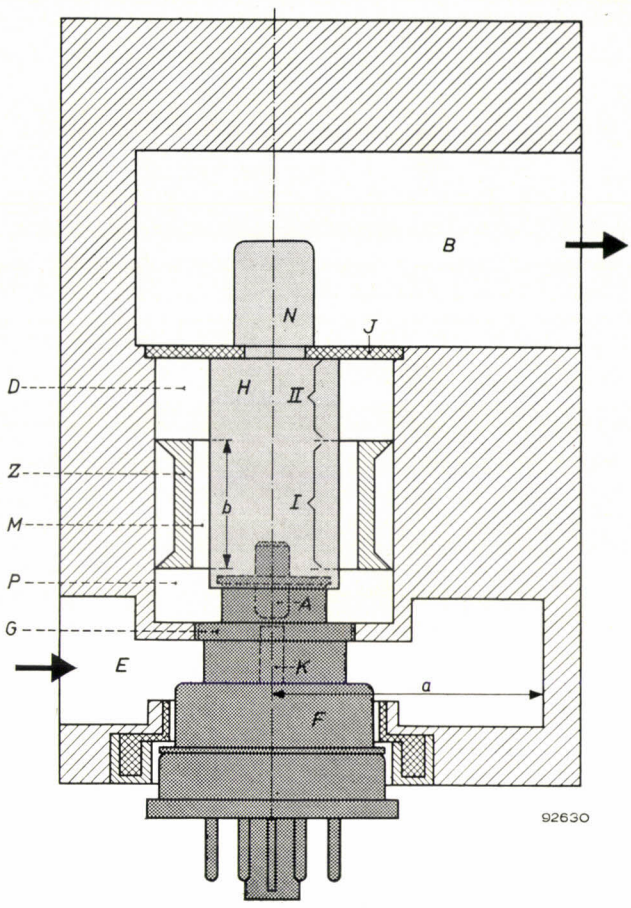


Fig. 1. Simplified cross-section of the amplifier. This consists of a block of brass, in which slots are cut to form the input and output waveguides, E and B respectively. Between these waveguides there is a cylindrical hole D. G, F, K and A are respectively the grid disc, the cathode disc, the cathode base and the anode of the disc-seal triode (EC 56 or EC 57). In cavity D there is an inner conductor H, which is supported by a disc J of plastic and which, with an extension piece N, protrudes into waveguide B. The anode resonant cavity P is tuned by a plunger Z. M is the gap between H and Z. H, Z and the wall of cavity D form two coaxial lines, denoted by I and II. The arrows indicate the direction in which the input power enters and the output power leaves the amplifier.

Construction of the amplifier

Fig. 1 shows in schematic form the construction of the amplifier [4]). It consists of a block of brass into which slots have been cut to form the input and output waveguides, E and B respectively. The two waveguides are coupled by a cylindrical hole D. The tube is inserted from underneath through the input waveguide, the grid disc G being screwed in the upper wall of E. The cathode disc F is insulated for DC from the underside of the input waveguide. For the operational frequency range, however, the coupling between these elements is practically a short-circuit [5]). The cathode base K protrudes through the input waveguide, which is short-circuited at a distance a of about one-quarter wavelength from the centre of the tube.

In the cylindrical cavity D there is an inner conductor H, the top of which is supported by a disc J of a synthetic material (polytetrafluorethylene) while the bottom makes spring contact with the anode disc. The cavity D also contains a plunger Z, which is in contact with the wall of D but not with the inner conductor H. The plunger thus forms with the part of H which it covers a coaxial line whose length b is approximately one-quarter of a wavelength. We shall refer to this coaxial line as I. Since the gap M between Z and H is fairly narrow, the characteristic impedance of I is low.

The section of the inner conductor above the plunger forms, with the wall of D, another coaxial line, which we shall refer to as II. Its characteristic impedance is higher than that of I. The inner conductor H has an extension piece N which protrudes into the output waveguide B. The dimensions of this section are so chosen as to make its radiation resistance over a wide frequency band equal to the characteristic impedance of coaxial line II. If waveguide B has a matched termination, the same will apply to the top end of coaxial line II.

At the bottom end of the plunger a resonant cavity P is formed, which incorporates the grid-anode capacitance of the tube. This cavity is inter-

[4]) A similar amplifier has already been described by A. E. Bowen and W. W. Mumford, A new microwave triode: its performance as a modulator and as an amplifier, Bell Syst. tech. J. **29**, 531-552, 1950.

[5]) A coupling of this kind is known as a "choke coupling"; see e.g. G. L. Ragan, Microwave transmission circuits, Radiation Laboratory Series, No. 9, McGraw-Hill, New York 1948.

rupted at the top by the bottom of gap M. Coaxial line I now functions as a quarter-wavelength transformer, the top end of which is terminated by the characteristic impedance of coaxial line II. The input impedance at the bottom end of this transformer is therefore very low (approximately 0.5 Ω), and it is this that forms the load of the anode resonant cavity P. The cavity is tuned by moving the plunger Z. (Owing to the matched termination of coaxial line II, the tuning does not alter the load resistance of the anode resonant cavity.)

The anode voltage is fed to the tube via a thin wire (see Y in fig. 13) which runs from N to the wall of the waveguide as far as possible at right angles to the electrical lines of force; it is led out via a disc capacitor mounted in a coaxial plug.

The input circuit

At a frequency of 4000 Mc/s the input impedance of the tube in the amplifier is almost a pure resistance, the value of which roughly corresponds to that needed for the waveguide E to be matched.

This matched termination gives rise to a standing-wave ratio of 1 in the waveguide. With a given tube, however, this is not readily possible for all frequencies in the desired frequency band. By way of example, *fig. 2* shows the standing-wave ratio in the waveguide E plotted as a function of the frequency. It can be seen from this figure that the input of the tube behaves rather like a tuned circuit.

At low frequencies the input resistance of a triode with grounded grid is approximately equal to the reciprocal value of the transconductance. In the case of the EC 57, the transconductance is approximately 20 mA/V, and hence the input resistance should be 50 Ω. As explained in the article quoted under [1]) the transconductance at 4000 Mc/s is about 30% lower than at low frequency, in which case therefore the input resistance should be 70 Ω. It must be taken into account that this input resistance is directly between grid and cathode.

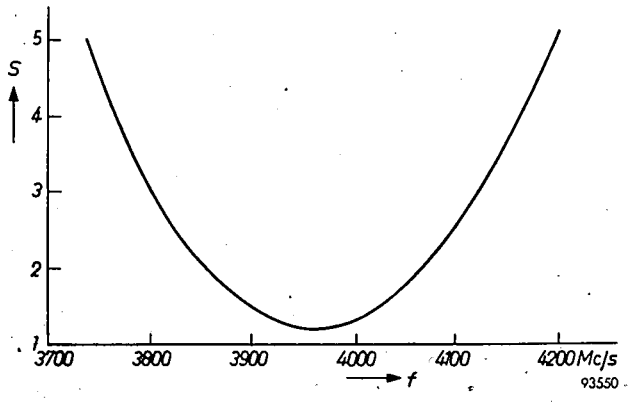


Fig. 2. Standing-wave ratio S in the input waveguide as a function of the frequency f.

At the high frequencies in question the equivalent resistance of the tube between cathode and grid disc has a different value, owing to the influence of the self-inductances and mutual capacitances of the electrodes. An equivalent circuit for the input impedance of the tube is shown in *fig. 3*. Here, R_{gk} is the input resistance, measured directly between grid and cathode and C_{gk} is the capacitance between these electrodes, while L_k is the inductance of the cathode base and the cathode disc. In parallel to these is the capacitance C_d between grid disc and cathode disc. A simple calculation reveals that at a frequency of 4000 Mc/s the equivalent resistance of this circuit is approximately a pure resistance of 400 Ω, which more or less corresponds to the characteristic impedance of the waveguide.

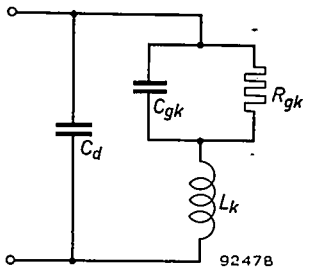


Fig. 3. Equivalent circuit of the amplifier input. R_{gk} is the input resistance between grid and cathode, C_{gk} the capacitance between these electrodes; L_k the inductance of cathode base and cathode disc, and C_d the capacitance between grid disc and cathode disc.

In order to be able, with individual tubes, to match the input end of the tube to the waveguide at any desired frequency in the frequency band to be covered, variable reactive elements have been introduced at two different points in the waveguide; these are not shown in fig. 1, but can be seen separately in *fig. 4a*. Each of these elements consists of a screw S in the centre of the upper wall of the waveguide, and a rod T joining the upper and lower walls, away from the centre. Waveguide E contains two such combinations at a distance d of approximately one-eighth of a wavelength apart.

It can be shown with the aid of the theory of electromagnetic wave propagation in waveguides that a rod, as represented by T, has the same effect on these waves as a coil, mounted between the two conductors of a Lecher line, would have on wave progagation along the latter. The inductance of this equivalent coil depends upon the position and the thickness of the rod. On the other hand, a screw S (fig. 4*a*) is equivalent to a capacitor connected between the conductors of a Lecher line. The capacitance of S depends upon the depth to which it protrudes into the waveguide. The combination of S and T is comparable, then, with a resonant circuit between the conductors of a Lecher line. Fig. 4*b* shows an electrically equivalent circuit, in which the waveguide is replaced by a Lecher line. The two capacitors (i.e. the screws S) can be so adjusted as to make the circuits either capacitive or inductive, and it is possible in this way, with varying tube impedances and with different frequencies, to match the waveguide E perfectly.

Although the two reactive elements make it possible to match the tube to the input waveguide E under all practical conditions, they nevertheless reduce the bandwidth in which a sufficiently small reflection appears at each side of the tuned frequency. This bandwidth, however, is still much larger than that of the anode circuit, and therefore the latter mainly determines the total bandwidth of the amplifier.

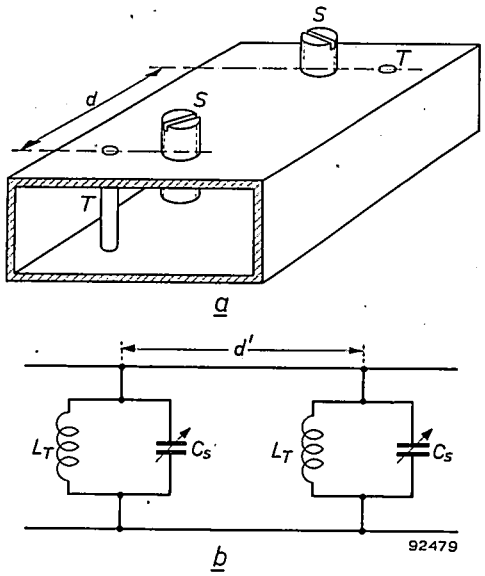


Fig. 4. *a*) The input waveguide contains two variable reactive elements, each of which consists of a screw S and a rod T. The distance d is approximately one-eighth wavelength. *b*) Equivalent circuit of the arrangement in fig. 4*a*. C_S and L_T are respectively the "capacitance" of the screw S and the "inductance" of the rods T. The distance d' is about one-eighth wavelength.

The anode circuit

In order to make the frequency characteristic flat enough over a wide frequency band, substantial damping of the resonant cavity P (see fig. 1) would be necessary. This would reduce the resonance resistance to the extent that no gain would be obtained. A familiar method of obtaining a sufficiently high gain over a wide frequency band at lower frequencies is to employ a band-pass filter, consisting of two resonant coupled circuits. It would be rather difficult mechanically to arrange a second resonant cavity near to P and directly coupled with it, particularly since both cavities would have to be tunable. In very-high frequency technique, however, it is possible to couple two resonant circuits at some distance apart by means of a suitable length of transmission line, which may consist of a coaxial line, a Lecher line, a waveguide or a combination of these. The latter is used in the present case. The second resonant circuit is contained in waveguide B, and consists of an inductance formed by two rods T_i joining the upper and lower walls, and a capacitance formed by a screw S_i through the upper wall (see *fig. 5*). This circuit, known as a resonant iris, will be referred to in the following simply as "*iris*" [6]). The position of this iris is so chosen as to make the system between the top of the plunger S and the iris electrically equivalent to a line one quarter-wave length long.

An equivalent circuit for the whole anode system is given in *fig. 6a*, in which C_a and L_a represent the resonant cavity P. The tube is represented by a voltage source V_a in series with the circuit L_a, C_a. (The damping effect of the internal resistance of the tube on the resonant cavity can be disregarded here.) The resonant circuit formed by L_i and C_i represents the iris sketched in fig. 5. This circuit is damped by the characteristic impedance R_0 of waveguide B. The coupling between both circuits, which in reality consists of the coaxial line I (see fig. 1), the coaxial line II, the extension piece N of the inner conductor H and the output waveguide B, is represented in fig. 6*a* by a transformer Tr having a transformation ratio n [7]). We shall now proceed, with reference to this diagram, to calculate the voltage V_i when the frequency of V_a is varied. For this purpose we shall substitute the section of the circuit on the left of L_i, C_i by a voltage source $V_a' = V_a/n$ in series with a capacitor having a capacitance $C_a' = n^2 C_a$, and a coil having an inductance $L_a' = L_a/n^2$ (fig. 6*b*). The circuit L_a', C_a' is

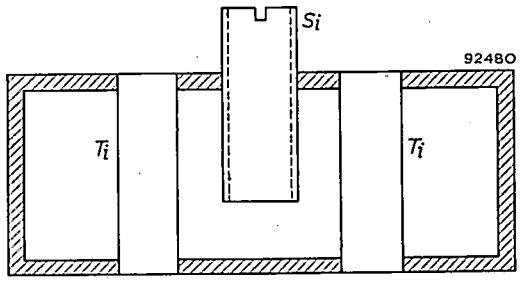


Fig. 5. The output waveguide contains a resonant iris, which consists of a screw S_i and two rods T_i.

[6]) The reason is that a reflecting element of this kind placed in a waveguide has the same effect as a conductive plate with a hole in it; see e.g. C. G. Montgomery, R. H. Dicke and E. M. Purcell, Principles of microwave circuits, Radiation Laboratory Series, No. 8, McGraw-Hill, New York 1948, Chapter 6.

[7]) This transformation ratio is dependent on the frequency; it appears, however, that sufficiently accurate results are obtained from a calculation in which n in the frequency band concerned is regarded as constant.

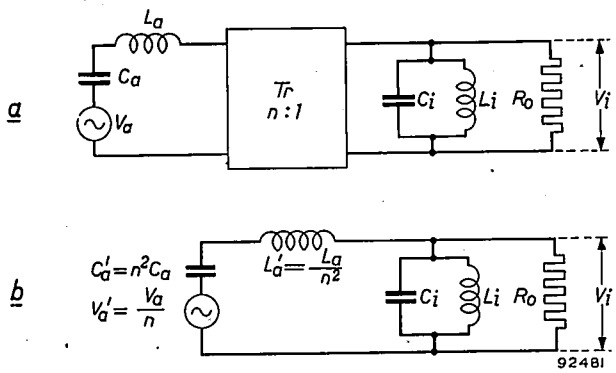


Fig. 6. *a*) Equivalent circuit in which the tube is represented by a voltage source V_a. L_a and C_a represent the anode resonant cavity, while L_i and C_i represent the iris. The iris is damped by the characteristic impedance R_0 of the output waveguide. The transmission system between resonant cavity and iris is represented by a transformer Tr with a transformation ratio n.
b) The section on the left of L, C_i is replaced by a voltage source V_a' in series with a capacitor C_a' and a coil L_a'.

then tuned to the same frequency as the circuit L_a, C_a, and hence the same frequency as L_i, C_i. The relevant angular frequency will be denoted ω_0. We shall also introduce the following quantities:

$Q_a = \dfrac{\omega_0 L_a'}{R_0}$, the figure of merit of circuit L_a', C_a' if the circuit L_i, C_i is not present. Evidently, this is also the figure of merit of the anode resonant cavity without the iris.

$Q_i = \dfrac{R_0}{\omega_0 L_i}$, the figure of merit of circuit L_i, C_i if this is damped by R_0.

$k = \dfrac{Q_i}{Q_a}$, the coupling factor, the meaning of which will be discussed later.

$\beta = \dfrac{\omega}{\omega_0} - \dfrac{\omega_0}{\omega}$. If the difference $\Delta\omega$ between ω and ω_0 is small with respect to ω_0, we may write $\beta = 2\Delta\omega/\omega_0$.

For the ratio between the voltages V_i and V_a' the following equation may now be derived:

$$\frac{V_i}{V_a'} = \{1 + (\beta Q_a)^2(1-2k) + (\beta Q_a)^4 k^2\}^{-1/2}. \quad (1)$$

In *fig. 7*, V_i/V_a' is plotted as a function of βQ_a for several values of k. For large values of k the curves show two peaks, for small values only one. The transition between these two situations takes place when

$$k = \tfrac{1}{2}. \quad (2)$$

The quantity Q_a is dependent on the transformation ratio n and hence of the strength of the coupling between the resonant cavity and the rest of the circuit. A small value of n is associated with a large value of L_a', and thus also with a large value of Q_a. In this case, then, k is small, and the resonant cavity is loosely coupled to the remaining part of the anode circuit. For this reason the quantity k is termed the coupling factor. The situation where $k = \frac{1}{2}$ is known as "transitional coupling" [8]).

As appears from fig. 7, the curve is flattest at approximately the value of k given by (2); it is therefore desirable to design the anode circuit so as to approximate to the transitional coupling. According to the definition of k, the figure of merit of the iris should be half that of the anode resonant cavity.

[8]) See e.g. C. B. Aiken, Two-mesh tuned coupled circuit filters, Proc. Inst. Rad. Engrs. **25**, 230-272, 1937. This term should not be confused with the term critical coupling which, with a band-pass filter both of whose circuits are equally damped, designates the coupling which gives maximum gain and for which the resonance curve is only just short of having two peaks. Although the transitional coupling in the circuit of fig. 6*b* corresponds to the latter situation, in this case the maximum value of V_i/V_a' does not occur. For $\beta = 0$ it follows from formula (1) that $V_i/V_a' = nV_i/V_a = 1$. We see, then, that V_i/V_a' is independent of n, while V_i/V_a is inversely proportional to n.

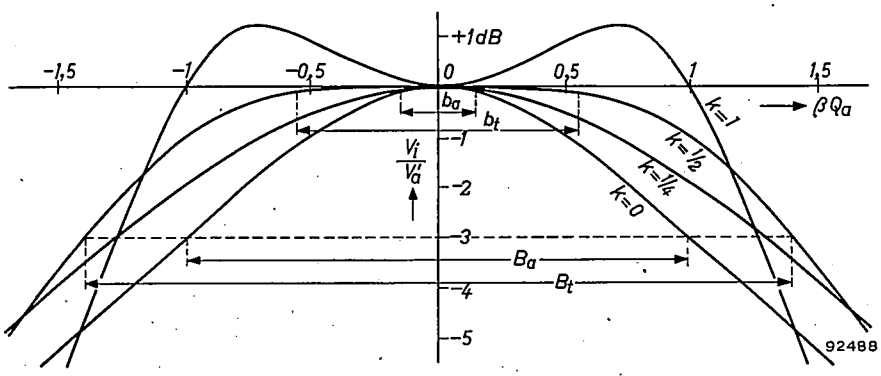


Fig. 7. The voltage ratio V_i/V_a' from fig. 6*b*, plotted in dB as a function of the product βQ_a. The curves hold for three different values of k. At $k = \frac{1}{2}$, transitional coupling is obtained.

The curve for $k = 0$ in fig. 7 indicates the value of V_i/V_a' for the case where no iris is present ($Q_i = 0$ means $L_i = \infty$, hence $C_i = 0$). This, then, is the resonance curve of the anode cavity separately.

The usual method of measuring the figure of merit of tuned circuits is to determine the width of the frequency band at the limits of which the circuit impedance is a factor $\sqrt{2}$ smaller than the resonance resistance. At the edges of this frequency band, the response is 3 dB down. This bandwidth of the resonant cavity is denoted in fig. 7 by B_a. If the corresponding bandwidth of the iris be denoted by B_i, it follows from the fact that the bandwidth of the circuit is inversely proportional to the figure of merit that:

$$B_i = 2B_a. \quad \ldots\ldots \quad (3)$$

In fig. 7, B_t indicates the width of the frequency band at the edges of which the curve for $k = \frac{1}{2}$ has fallen to 3 dB below peak value. It is seen that $B_t = \sqrt{2}B_a$. The improvement obtained by applying the iris is, however, much larger than this factor $\sqrt{2}$. The reason is that in order to keep the distortion of the transmitted signals within reasonable limits, a much smaller frequency band can be used than is indicated by B_t in fig. 7. As a rule, the requirement is that the amplification at the edges of the frequency band should drop by no more than 0.1 dB. In the following we shall denote this bandwidth by a small letter b in order to distinguish it from the earlier-mentioned bandwidth B. It will be clear from a consideration of fig. 7 that in this respect the curve for $k = \frac{1}{2}$ is appreciably better than that for $k = 0$. For a single tuned circuit the bandwidth b_a in question is

$$b_a = 0.15\ B_a.$$

If transitional coupling is now introduced between this circuit and a second, whose bandwidth is $2B_a$, the resultant curve will fall to 0.1 dB below the peak value at a width b_t which is given by

$$b_t = 0.39\ B_t = 0.55\ B_a.$$

Thus, by introducing the iris, the width of the frequency band at the edges of which the amplification has dropped by only 0.1 dB is increased by a factor $0.55/0.15 = 3.7$.

Adjustment to the required bandwidth

When the frequency band b_t of an amplifier constructed according to the foregoing principle is required to have a specific value, it is easy to calculate what values of the bandwidths B_a and B_i of the anode resonant cavity and the iris are required, since

$$B_a = \frac{1}{0.55}\, b_t = 1.82\ b_t \text{ and } B_i = 2\ B_a = 3.64\ b_t.$$

If, for example, constant amplification is required over a frequency band of 55 Mc/s, the bandwidth B_a of the amplifier without iris should be $1.82 \times 55 = 100$ Mc/s and the bandwidth B_i of the iris $3.64 \times 55 = 200$ Mc/s. In order to adjust the latter band width tothe required values an experiment is needed to determine the position and thickness of the rods T_i (see fig. 5).

The iris is placed in a waveguide, one end of which is matched and the other end connected to an energy source, also matched to the waveguide. The iris is then damped, both by the load and by the energy source, with an impedance equal to the characteristic impedance of the waveguide. An equivalent circuit, in which the impedances are represented as lumped circuit elements, is shown in *fig. 8*. The circuit L_i, C_i is now damped by a resistance $\frac{1}{2}R_0$, and it is thus evident that the circuit must be so designed that its bandwidth in this arrangement will be equal to $2B_i = 7.3\ b_t$, in our case 400 Mc/s. This situation can be achieved by a suitable choice of the position and thickness of the rods T_i (fig. 5).

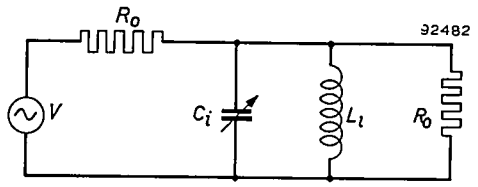


Fig. 8. Equivalent circuit for the iris, placed in a waveguide with matched termination; the waveguide is connected to an energy source which is also matched.

Once the rods have been placed in the amplifier with the correct thickness and at the correct place, the resonant cavity and iris must be tuned to the required central frequency, and the coupling adjusted such that $k = \frac{1}{2}$. The resonant cavity is tuned with the plunger Z (fig. 1) and the iris with the screw S_i (fig. 5). To adjust the coupling it is necessary to vary the figure of merit Q_a of the resonant cavity. This is done by changing the transformation ratio n (see fig. 6). One way might be to vary the characteristic impedance of the quarter-wavelength transformer by altering the radial dimensions of gap M (see fig. 1). Mechanically, this would be difficult; however, for small variations, the same effect can be achieved with a construction in which a reactive element is placed in the waveguide at a distance of one-eighth of a wavelength from the iris. This element again consists of a rod joining the upper and lower walls of the waveguide and of

a screw through the upper wall. In *fig. 9* a cross-section of the output waveguide is shown in which this reactive element is represented by a screw S'

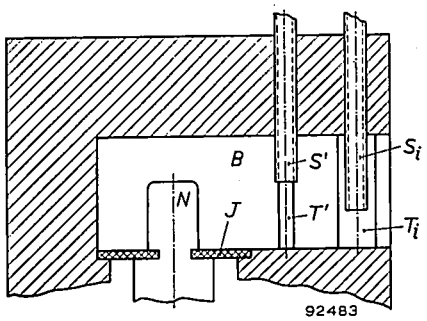


Fig. 9. Cross-section of the output guide, in which the iris is denoted by S_i, T_i, and the reactive element with which the coupling is adjusted is denoted by S', T'. For the meaning of letters B, N and J see fig. 1.

and a rod T'. Now, it can be shown that the effect of screwing S' in and out is to vary the transformation ratio n of the transmission system between resonant cavity and iris, and also to change V_a', L_a' and hence Q_a. Since Q_i remains constant, the coupling factor k is also varied during this process.

Cooling

In the construction of the amplifier as shown in fig. 1 the anode section of the tube is entirely incorporated in the inner conductor H. There is therefore a danger that the anode disc will be overheated at the maximum anode dissipation of 10 W and for this reason, measures have been taken to cool the inner conductor H. For this purpose, H could be made hollow and provided internally with radial cooling fins. *Fig. 11* shows a cross-section of such an inner conductor, in which the direction of air flow is indicated by arrows. Through a polystyrene tube U the cooling air is fed in from outside to the extension piece N of the inner conductor H. The air flows via the cooling fins V to the output waveguide B, where it leaves the amplifier. The quantity of air required is about 6 l/min at a pressure of approximately 5 cm water.

The complete amplifier

Fig. 12 is a photograph of an amplifier as seen from the input and output sides; *fig. 13* gives a perspective cut-away sketch of the complete

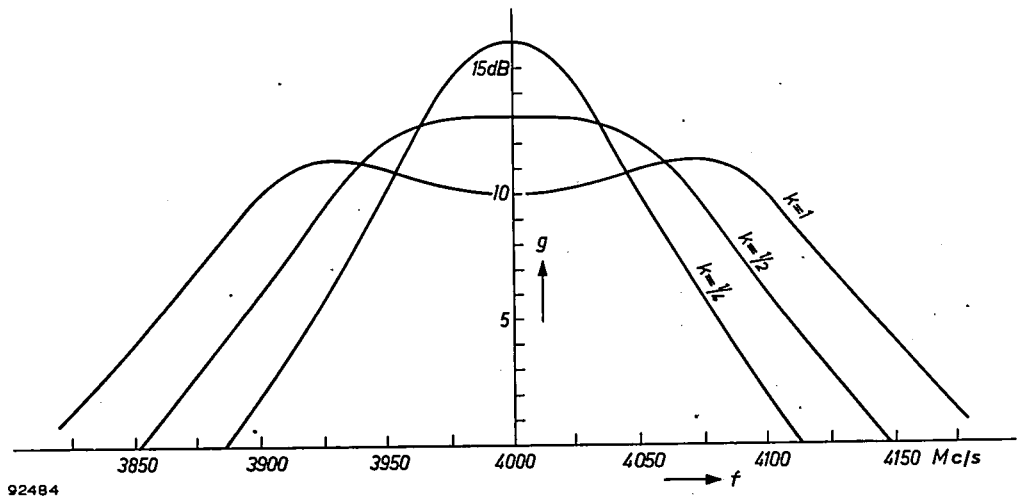


Fig. 10. The gain obtained with an tube EC 57, tuned to a central frequency of 4000 Mc/s, for various values of the coupling factor k, and an iris bandwidth B_i of 200 Mc/s.

As regards the curves in fig. 7, this affects both the two scales and the running parameter k, so that the total effect of turning screw S' is difficult to see from this figure. A good impression of what happens when the coupling screw S' is turned is given by *fig. 10*, in which, for an EC 57 tube, the amplification in dB is plotted against the frequency for several values of k. It can be seen that as k decreases, the amplification at the central frequency increases, although the bandwidth becomes smaller. At the transitional coupling ($k = \frac{1}{2}$) the amplification at the central frequency is 13 dB.

amplifier showing the control elements discussed in the foregoing.

The plunger Z is moved by two rods W fixed at pivots to a cross-beam X, forming, as it were, a balance. In this way the plunger is always subjected to two equal forces, thereby precluding mis-alignment.

The EC 57 tube [1]) can be used in this amplifier. If a smaller output power is adequate, a type EC 56 tube may be used instead, the nominal anode current of which is half that of the EC 57 (30 against 60 mA). This current is set to the correct value by

means of a variable cathode resistor; to reduce the effect of deviations in the characteristics of individual tubes, the cathode resistor can be connected to a point at a negative potential.

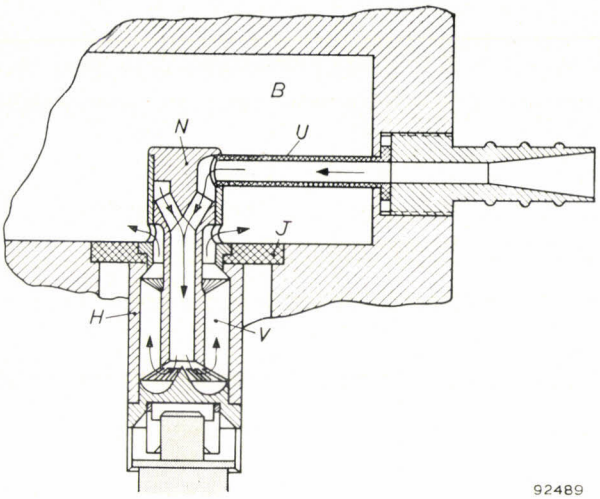


Fig. 11. A system for cooling the amplifier. The cooling air is fed through the polystyrene tube U to the cooling fins V. For the meaning of the letters B, H, J and N see fig. 1.

The gain obtainable in the case of small signals is 12 dB with the EC 56 and 13 dB with the EC 57 (at a bandwidth b_t of 55 Mc/s); for larger signals the gain drops. In *fig. 14* the gain for both tubes is plotted as a function of the output power. It can be seen that with a gain of 8 dB an output of 0.5 W can be obtained from the EC 56, while with the same gain the EC 57 delivers a power of 1.5 W. Fig. 14 holds for an amplifier with a bandwidth b_t adjusted to 55 Mc/s. As mentioned in the article quoted under [1]), a higher gain can be obtained with a smaller bandwidth, in such a way that the product of bandwidth and gain is almost constant. The bandwidth cannot, however, be reduced too far: at a bandwidth of 10 to 20 Mc/s the amplifier becomes unstable owing to the feedback occurring in the tube.

Feedback

When the tube is in operation, feedback occurs owing to the small but by no means negligible capacitance between anode and cathode. The reason for this is that the grid does not completely screen these two electrodes from each other. A second cause of feedback is the fact that, owing to the electromagnetic alternating field in the anode resonant cavity, currents are induced in the grid wires, and these currents in their turn give rise to an alternating voltage between grid and cathode. The effect of this can be accounted for in the equivalent circuit by a coil in series with the grid connection [9]).

[9]) See G. Diemer, Passive feedback admittance of disc-seal triodes, Philips Res. Rep. **5**, 423-434, 1950.

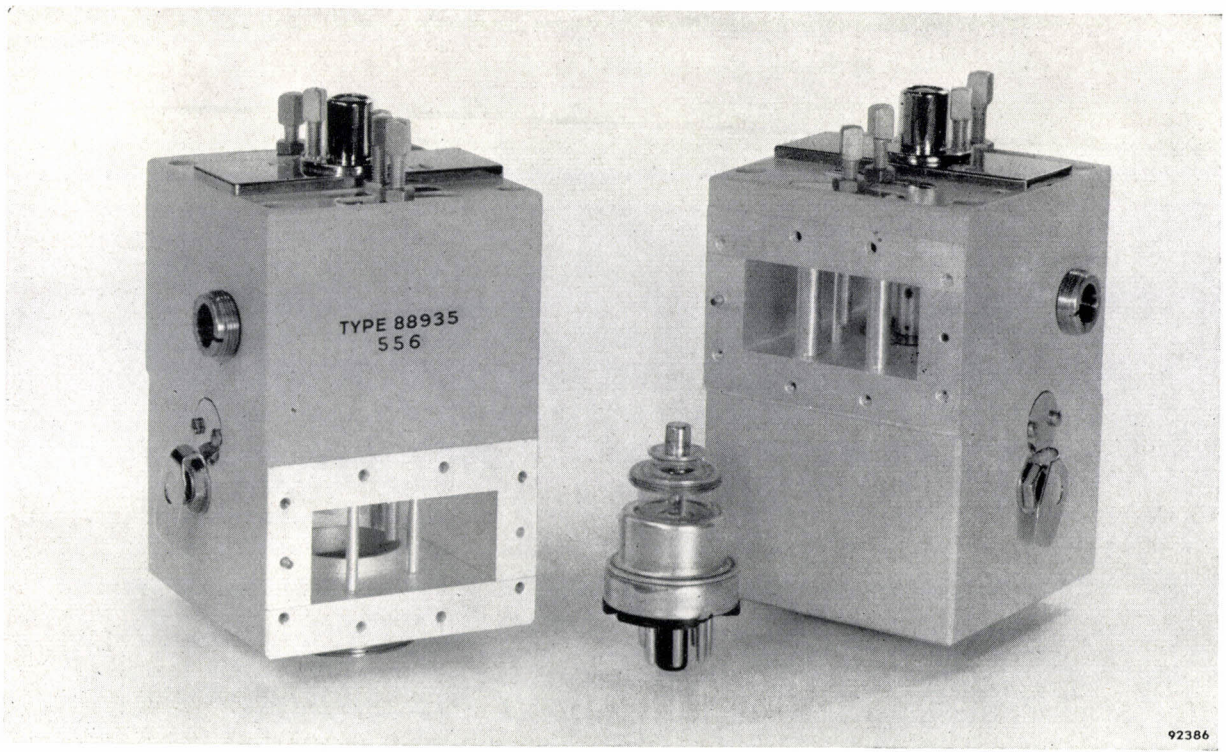


Fig. 12. Two amplifier blocks, the left one seen from the input side, the right one from the output side. The EC 57 tube is shown between them.

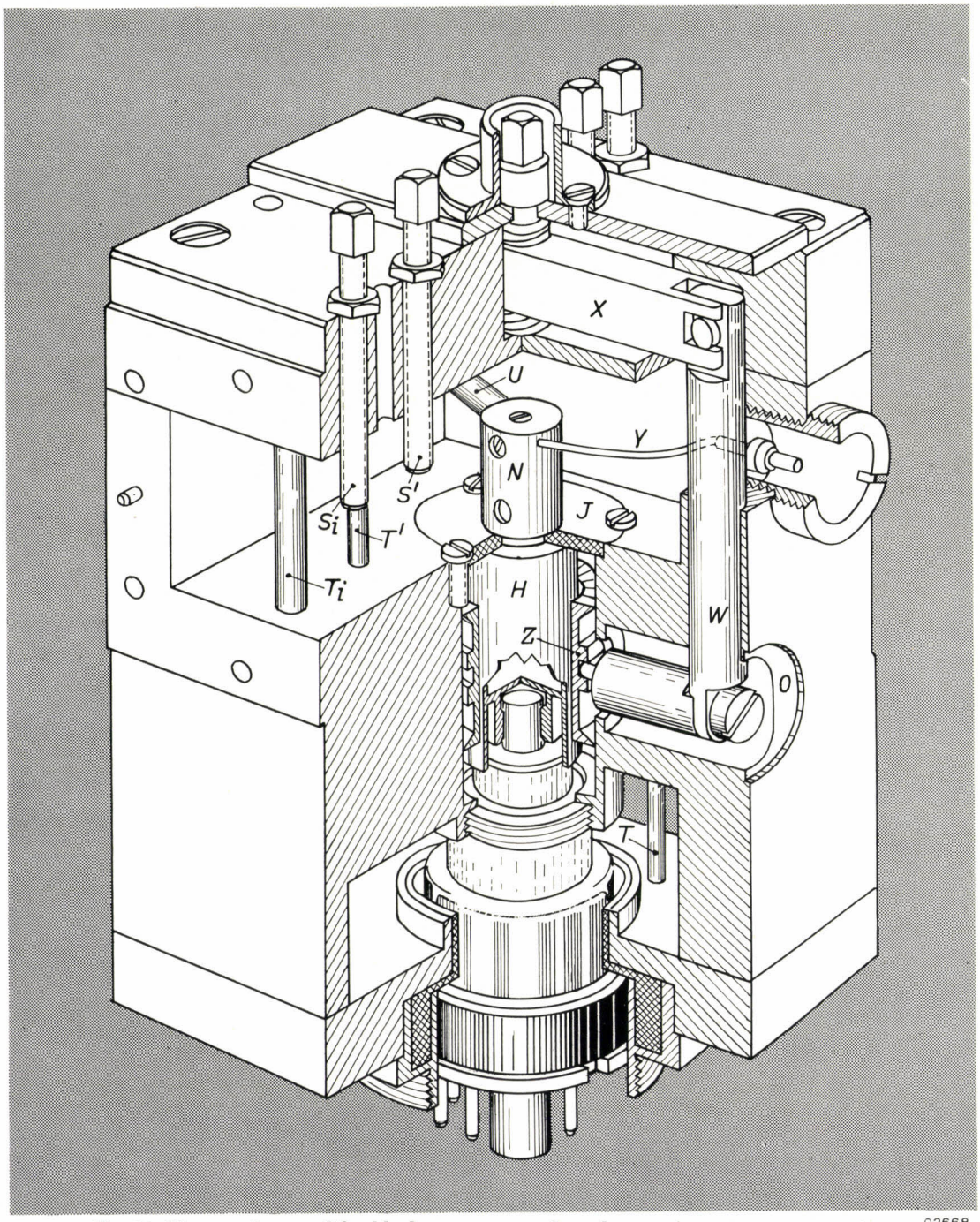


Fig. 13. The complete amplifier block, cut away to show the most important components. W denotes the driving rods for plunger Z. These rods are fixed at pivots to cross-beam X. Underneath they carry pins which engage with the plunger. The other letters have the same meaning as in earlier figures.

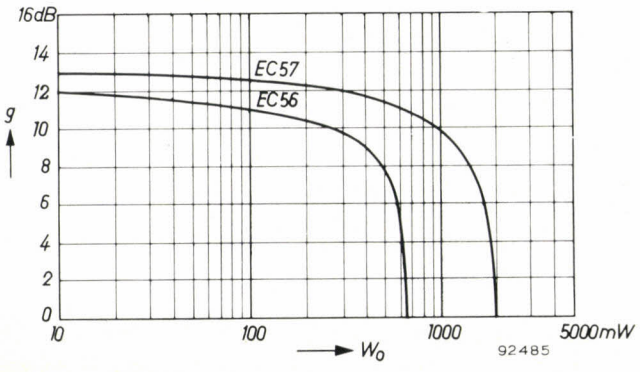


Fig. 14. Gain g as a function of the output power W_0 for tubes EC 56 and EC 57, at a bandwidth b_t of 55 Mc/s.

The two causes of feedback oppose each other and at a given frequency, can even cancel each other out. To illustrate this, an equivalent circuit is shown in *fig. 15a* for a tube in which only the heater voltage is switched on. The elements C_{ak} and L_g are responsible for the two feedback effects. Applying a delta-star transformation to the three capa-

citors, we obtain the result as shown in fig. 15*b*. It will be evident that no feedback appears at the resonant frequency of L_g and the equivalent capacitance C_3 in series. With the tubes EC 56 and EC 57 this resonant frequency is at present about 5200 Mc/s and is thus considerably higher than the frequencies at which the amplifier will be used. At a frequency of 4000 Mc/s the series arrangement of L_g and C_3 therefore has a reactance 2.5 times smaller than the reactance of C_3 alone. It may broadly be said, then, that owing to the presence of L_g the capacitance C_{ak} is apparently reduced by the same factor.

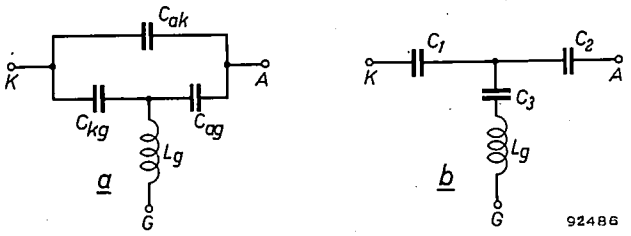


Fig. 15. *a*) Equivalent circuit for a tube with only its heater voltage applied. *A* anode, *K* cathode, *G* grid, L_g equivalent inductance of the grid wires.
b) Diagram obtained by applying a delta-star transformation to the three capacitors.

When the tube is in operation the situation is considerably more complicated, mainly because of the effects of space charge and electron transit time, which increase the input and output damping and also give rise to a phase shift between the voltages on cathode and anode. No complete theory has yet been presented to account for all these facts, but enough is now known to explain some of the phenomena occurring in a tube under operating conditions.

The feedback in the tube need not be regarded entirely as a drawback. As pointed out by Van der Ziel and Knol [10]), feedback can be used to increase the product of gain and bandwidth. In fact, the favourable properties of tubes, EC 56 and EC 57 are partly due to feedback.

In another respect, however, feedback does cause difficulties, expecially if one wishes to connect several of these amplifiers in cascade. It is then obviously convenient to arrange for the anode circuit of each tube to form a band-pass filter with the cathode circuit of the next tube. When this is done the role of the iris, which we discussed when dealing with the anode circuit, is taken over by the cathode circuit of the next tube. Owing to the feedback, however, the cathode circuit of the tube can no longer be regarded as a simple resonant circuit. For instance, if the input side of a tube is matched to the input waveguide while the anode resonant circuit is detuned, the tuning of this cavity can seriously upset the matching at the input side. Since the bandwidth of the anode circuit is much smaller than that of the cathode circuit, this interference occurs in a relatively narrow frequency band, so that it becomes very difficult to form with this input circuit a band-pass filter with transitional coupling. The effect is stronger the higher the resonance resistance of the anode resonant cavity and thus the narrower the bandwidth of the anode circuit.

In spite of these difficulties it has proved possible to connect three amplifier stages in cascade and to make the overall frequency characteristic sufficiently flat. One can never be certain, however, that the stages separately have a flat frequency characteristic, and there is accordingly a considerable risk that a change in the characteristics of the tubes will be associated with an unwanted distortion of the frequency characteristic.

The use of several stages in cascade can be made much easier by effecting the coupling between the stages via ferrite directional isolators.

The use of ferrite isolators

For an explanation of how a ferrite isolator functions reference is made to an earlier article in this Review [11]). The article describes, among other things, an isolator developed for use at frequencies of about 4000 Mc/s. By placing an isolator of this kind, both ends of which are matched to the waveguide, between two amplifier stages, many difficulties arising from feedback are overcome and, moreover, several other advantages obtained. Each stage can now, with its associated isolator, be tuned as a separate unit, and several such stages can readily be connected in cascade. The sacrifice this entails is a slight loss of power, amounting to less than 1 dB per isolator.

Fig. 16 shows three cascaded amplifier stages, coupled via ferrite isolators. Using two EC 56 tubes and one EC 57, the output power of the amplifier is 1.5 W with an input power of 1.5 mW. The total gain is thus 30 dB. With a bandwidth b_t per stage of 55 Mc/s, the bandwidth of the total amplifier is 42 Mc/s. If a tube has to be changed, it is only necessary to readjust the stage in question.

[10]) A. van der Ziel and K. S. Knol, On the power gain and the bandwidth of feedback amplifier stages, Philips Res. Rep. **4**, 168-178, 1949.

[11]) H. G. Beljers, The application of ferroxcube in unidirectional waveguides and its bearing on the principle of reciprocity, Philips tech. Rev. **18**, 158-166, 1956/57 (No. 6).

When an isolator is employed, the band-pass filter in the anode circuit of the tube is not (as with direct inter-stage coupling) partly formed by the cathode circuit of the following tube, but by the anode resonant cavity and a waveguide iris, in the manner already described. The input capacitance of the tube then has no effect on the properties of this band-pass filter, and since this capacitance is the quantity that varies most during operation, the shape of the frequency characteristic is better maintained when isolators are used.

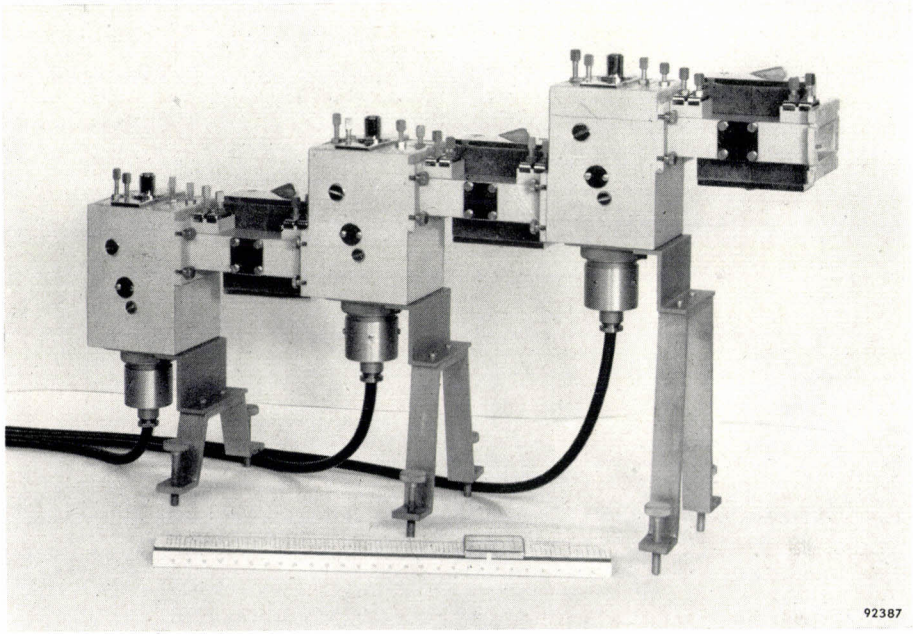


Fig. 16. Amplifier, consisting of three high-frequency stages coupled via ferrite isolators.

Group delay

For the undistorted transmission of frequency-modulated signals it is not only necessary that the amplification is constant over the whole frequency band concerned; it is equally necessary that the group delay in that band should be constant. Measurements and calculations have shown that a high-frequency amplifier, as described here, has much more favourable properties in this respect than a normal intermediate frequency amplifier, as used in most microwave links [12]). This is largely due to the fact that the absolute bandwidth of an IF amplifier is always much smaller than that of a high-frequency amplifier.

The group delay of an amplifier stage such as described above is 3.2×10^{-9} sec at the central frequency; at frequencies differing by 10 Mc/s from the central frequency, the group delay differs by only 64×10^{-12} sec from the above value. In consequence of these favourable figures it may, in certain circumstances, be advantageous for relay stations in microwave links to use no IF amplifiers but only a number of high-frequency amplifier stages according to the principle discussed. Since in this case the product of gain and bandwidth per stage is greater than in the case of an IF amplifier, it is possible to reduce the total number of stages in the amplifier, thereby considerably reducing the variation of group delay in the transmitted frequency band.

The above-mentioned figures for the group delay refer to a perfectly adjusted amplifier. It should be borne in mind, however, that after some time in operation the amplifier may no longer be exactly adjusted. (One cause may be a change in the inter-electrode capacitances of the tubes.) The usual consequence of this is a change in the shape both of the phase characteristic and the amplitude characteristic. The group delay is, of course, directly affected by deviations in the phase characteristic, but it can also be indirectly affected by changes in the amplitude characteristic, especially if the amplifier delivers a high power output, in which case the tube capacitances depend upon the magnitude of the signal voltages. As a result, an amplitude variation, caused in frequency-modulated signals by the amplitude characteristic having an unwanted

[12]) See J. P. M. Gieles, The measurement of group delay in triode amplifiers at 4000 Mc/s. This article will shortly appear in L'Onde Electrique.

shape, can be associated with a phase variation, which in turn gives rise to a change in the group delay. It has been found, however, that the latter effects are very small in the amplifier discussed in this article.

There are certain other less important causes of variations of group-delay in the amplifier, one of them being the electron transit time in the tubes and another the group-delay of the electromagnetic waves in the isolators. We shall not give details here of group-delay measurements carried out on this amplifier, but mention merely that the results of such measurements lead us to expect that it will be possible with 50 relay stations, equipped solely with amplifiers as discussed in this article, to transmit a colour television signal according to the NTSC system over a distance of 2500 km, without exceeding the maximum permissible distortion.

In the development of the amplifier described here, important contributions have been made by G. de Vries and A. Meyer of the Eindhoven Research Laboratory, H. Kramer of Philips Telecommunication Industry in Hilversum and W. Busscher, J. Jongsma and K. Rodenhuis of the Electronic Tubes Development Laboratory in Eindhoven.

Summary. This article describes the construction and properties of an amplifier for operation at frequencies from 3800 to 4200 Mc/s. The tube used in the amplifier may be an EC 56 or an EC 57, with which an output power of 0.5 and 1.5 W respectively can be obtained with a gain of 8 dB. The low level gain is 12 and 13 dB respectively. The frequency characteristic is flat so that the gain varies by no more than 0.1 dB in a frequency band of 55 Mc/s. The input and output of the amplifier are designed for connection to waveguides. In the anode circuit a band-pass filter is used, which is formed by a resonant cavity coupled with a resonant circuit in the output waveguide. The input and output waveguides contain variable reactive elements for input matching and for adjusting the anode circuit. When several of these amplifiers are connected in cascade it becomes difficult to adjust them correctly, owing to feedback. These difficulties are largely overcome by using ferrite isolators between the stages. The variation of group delay in the frequency band covered is much smaller in this amplifier than in the IF amplifiers normally used in microwave radio links. It may therefore be advantageous for relay stations to use no IF amplifiers but only a number of the high-frequency amplifier stages described here.

A LUMINOUS FRAME AROUND THE TELEVISION SCREEN

621.397.62

People generally prefer to view television in a lighted room. The brightness level in the room, around the set, however, is always considerably below that of the picture. An improvement in viewing conditions appears to be obtained by introducing a zone of intermediate brightness between the screen and its surroundings. Such a transition zone is generally provided by surrounding the picture with a light-coloured frame. The question now arises whether it might not be an improvement to provide this frame with its own illumination. Also, how wide and how bright should such a frame be? An answer to these questions was obtained by means of the following tests.

On a screen of 480 × 360 mm, which served as a substitute for a TV screen, no TV transmissions being available at the time of the tests, a film was projected by a 16 mm projector. The projector was fitted with a light-blue filter to make the colour of the projected image correspond to that of a TV picture. The projection screen was surrounded with a frame whose luminance and width could be varied with the aid of control knobs by an observer seated at a distance of 2.20 m from the screen (*fig. 1*). The luminance was variable between 0 and 34 cd/m²; the width between 0 and 160 mm. The surroundings, which were grey or black curtains, were given a known and adjustable luminance, uniform throughout the entire field of vision.

Whilst the observer watched and listened to the film, he was asked — at times when the average screen luminance, measured in advance, remained fairly constant for longish periods — to adjust the frame in accordance with:

a) the luminance he preferred at each of four given values of the frame width,

b) the width he preferred at each of four given values of frame luminance.

Graphical interpretation of the results made it possible to establish which combination of frame luminance and frame width the observer preferred for the various given values of the average screen luminance at three different luminance levels of the surroundings (see *Table*). Care was taken that

Table. Combinations of luminance values of the surroundings and average luminance values of the screen (in cd/m²) at which the adjustments of frame width and frame luminance were carried out.

Surroundings	0.005			0.5			5.0	
Screen	2.0	7.3	30	3.8	9.1	32	25	48

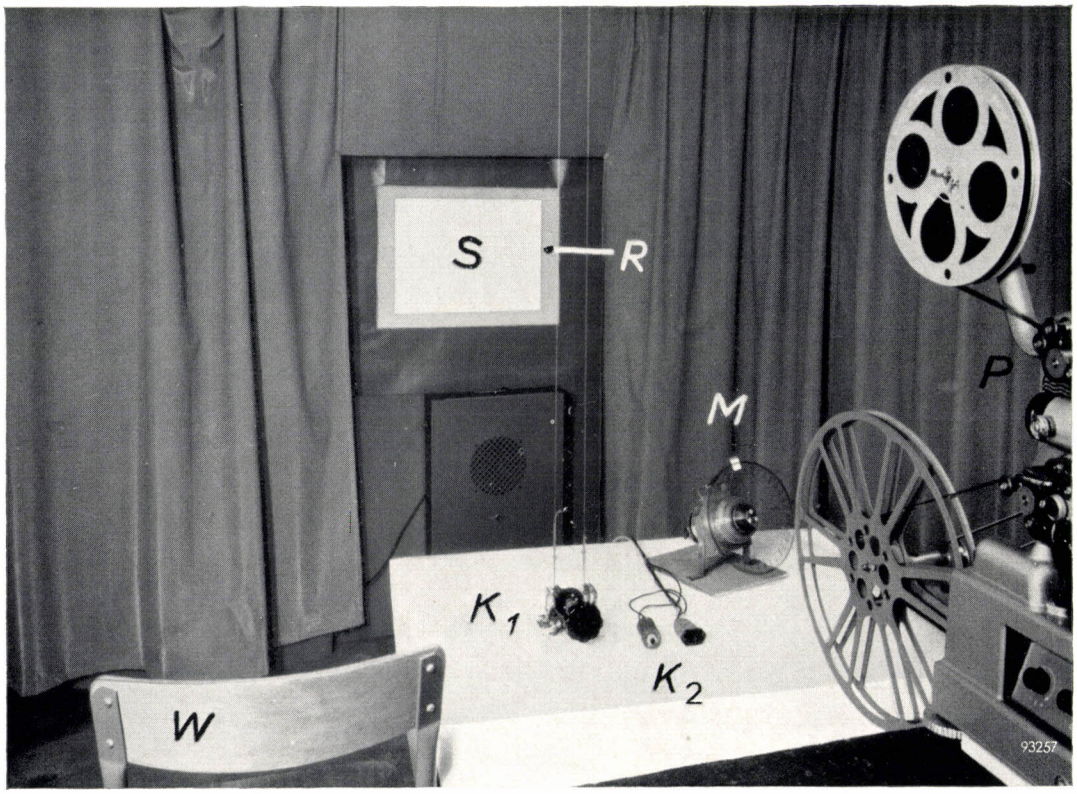


Fig. 1. Test arrangement for determining the most favourable width and brightness of the frame around a TV screen. A substitute for a TV picture is produced on the screen S by the film projector P. The observer, seated on chair W, can operate knob K_1 to adjust the luminance of the frame R around the screen, and the knobs K_2 (one being used for reversing the movement) for varying the width of this frame by means of two sliding masks. The ratio of the width of the frame along the short edge of the screen to that along the long edge remains constant at 4 : 3. The adjusted luminance can be read from disc M by the person conducting the test, who also measures the adjusted frame width. The desired luminance of the surroundings is obtained by a variable arrangement of uniform diffuse lighting and by variation of the reflection factor of the surroundings (other curtains).

no one observer saw the same film twice, in order to preclude above-normal attention to the frame due to lack of interest in the picture. The observer was explicitly told to carry out the adjustments whilst looking at the picture. For each set of luminance values of screen and surroundings (see Table), the adjustments mentioned under (*a*) and (*b*) were carried out by 20 to 25 observers. This provided us with the same number of diagrams of the kind shown in *fig. 2*, and the same number of points Q representing preferred combinations of frame width and frame luminance. *Fig. 3* shows the cluster of points obtained for a screen luminance of 25 cd/m^2 and a surroundings luminance of 5 cd/m^2. There is evidently a considerable difference between the individual preferences of the various observers.

The following conclusions can be drawn from the tests [1]). As regards the preferred frame *width*, no correlation was found with the luminance of the surroundings of the screen. The average preferred frame width over all the tests was 0.3 times the half screen dimension. This value lies close to the customary widths of frames on modern TV receivers.

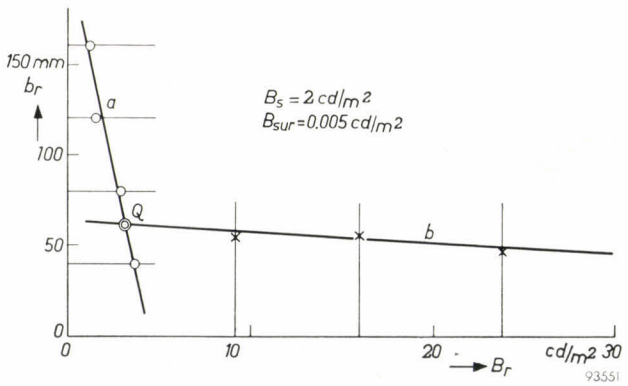


Fig. 2. For each of the four given frame widths b_r, the observer adjusted the frame luminance B_r to the level he preferred (circles), and for each of the four given frame luminances the same observer adjusted the width of the frame to the value he preferred (crosses). The intersection Q of line a (through the circles) and line b (through the crosses) represents the combination of frame luminance and frame width most preferred by this particular observer for the prevailing average luminance B_s of the screen and the luminance of the surroundings B_{sur}.

[1]) A more detailed account of these tests and of the statistical evaluation of the results is published in Light and lighting **50**, 245-250, 1957 (No. 8).

The preferred frame *luminance* was found to increase with the screen luminance as well as with the luminance of the surroundings (*fig. 4*). For modern TV sets in lighted rooms, it appears that a frame luminance somewhere between 10 and 20 cd/m² would be preferred. A statistical analysis of the test results further demonstrates that the diverging wishes of the large majority of viewers can be fulfilled by making the frame luminance variable between 0 and 40 cd/m².

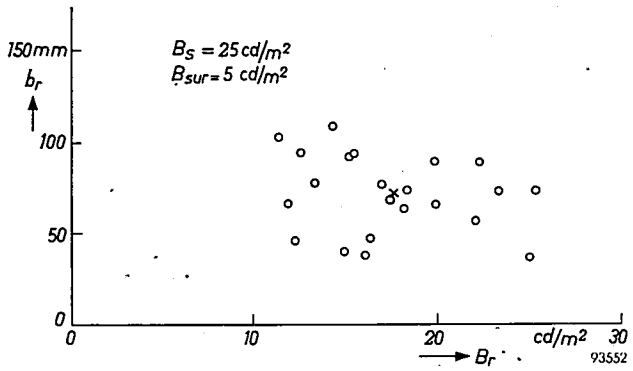


Fig. 3. Example of a cluster of points consisting of all points Q derived in the manner shown in fig. 2 (each observer supplying one point). This scatter diagram, which clearly illustrates individual differences, is, like fig. 2, only valid for a given set of luminance values of screen and surroundings. The cross shows the average preferred combination.

The customary non-illuminated screen frames derive their brightness from light reflected from the room and from the screen. Light-coloured frames might sometimes attain a luminance of 10 cd/m²;

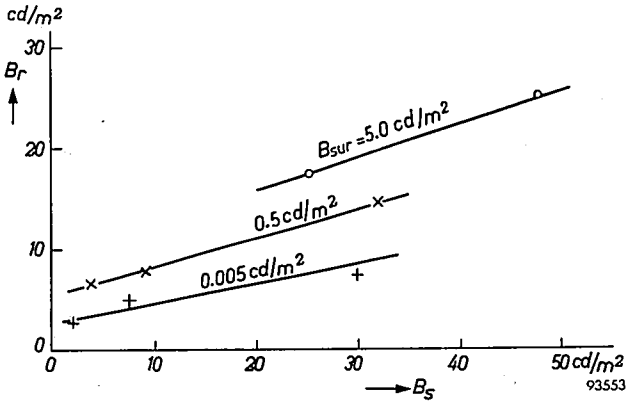


Fig. 4. Average values of the preferred frame luminances B_r as a function of the average screen luminance B_s, for three values of the luminance B_{sur} of the surroundings.

for a dull, fairly dark gray frame, 5 cd/m² is about the upper limit. We may conclude from this that separate illumination of the frame might sometimes be worth while.

As regards colour, the observers unanimously preferred a frame of the same colour as the picture.

J. J. BALDER.

LOW-VOLTAGE OSCILLOSCOPE TUBES

by F. de BOER and W. F. NIENHUIS.

621.317.755

The demand for cathode-ray oscilloscopes has greatly increased in the last decade. In addition to the demand for comprehensive instruments the need has arisen for small sets such as are employed in radio and television servicing workshops. It would be possible to effect a considerable reduction in the weight and bulk of oscilloscopes if only the components generating the voltage required for the final anode of the CRT could be made smaller. In this article a type of CRT is described which requires an anode voltage very much lower than usual, with a consequent reduction in the size and weight of the oscilloscope. Despite the low voltage on its anode the tube gives a good clear image and offers many possibilities; it will appeal to radio amateurs, amongst others.

Most oscilloscope tubes in common use require from 800 to 1500 volts on their anodes: one consequence of this was that all oscilloscopes have had to embody a heavy, bulky and expensive high tension transformer. An obvious first step towards building the desired small lightweight oscilloscopes was to try to design a CRT that would work satisfactorily with a low anode voltage. Lowering the anode voltage has the additional advantage that the sensitivity of the tube is increased, thus reducing the amplification that the test signal has to undergo; the appropriate circuits can then be simplified.

We have now succeeded in making such tubes. The DG 7-32 and DG 7-31, which will be described below, require a voltage of only 400 V (minimum).

The use of an anode voltage as low as this means that the composition of the fluorescent layer and the manner in which it is applied have to meet highly specialized requirements; the electrode assembly too has to be suited to the low anode voltage if good resolution (i.e. a small enough spot) is to be obtained.

The nature of these problems and the solutions found for them will be discussed below. A small cathode-ray oscillograph making use of a tube of this type is described briefly at the end of the article. It should be noted at this juncture that the DG 7-32 tube has been designed for a sweep voltage that is symmetrical with respect to earth; in the DG 7-31 tube one of the plates is earthed, and the sweep voltage is therefore applied asymmetrically. The two tubes are of the same size and have a screen diameter of 7 cm.

The fluorescent screen

For a tube working at these low anode voltages, the fluorescent screen cannot be fixed to the glass of the envelope with the aid of some soluble silicate binder [1]), as is the normal practice. Although the thickness of the film of binder lying over the screen is only of the order of 0.01 micron, electrons of about 400 eV lose more than 95% of their kinetic energy in passing through, with the result that the brightness of the trace is cut down excessively. Again, any chemical action on the grains of the fluorescent powder, such as might take place during the application of the screen, would be disastrous for the same reason.

This being so, the old and familiar "flow coating" method is the one preferred. A certain quantity of a suspension of the fluorescent powder in a solution of nitrocellulose in some organic solvent (e.g. ethyl acetate) is poured into the envelope. By rotating the tube the entire screen surface is covered with the liquid and then allowed to dry off. Subsequently the nitrocellulose is removed by heating up to 400 °C. Owing to the volatility of the decomposition products and the moderate temperature the fluorescent powder undergoes no chemical change.

It is well known that a layer of this kind cannot be used in conjunction with a low anode voltage without provision being made for the fact that the electron beam is not adequately balanced by secondary emission from the screen, which consequently acquires a charge [2]). Moreover, covering the fluorescent layer with either a conductive layer or a layer having a suitable coefficient of secondary emission (as is done in television picture tubes) is out of the question for the reasons already stated. However, the screen can be prevented from charging up by previously giving the screen portion of the envelope interior a transparent conductive

[1]) F. de Boer and H. Emmens, Philips tech. Rev. **16**, 232-236, 1954/55.

[2]) Cf. J. de Gier, A. C. Kleisma and J. Peper, Philips tech. Rev. **16**, 26-32, 1954/55.

coating that serves as base for the fluorescent powder. Tin oxide with certain additives is a suitable substance for the coating, conferring on it the properties of a semiconductor [3]). The coating is applied by spraying a tin chloride solution into the tube, which is heated to temperature approaching the softening point. The conductive coating only "takes" where the temperature of the glass is sufficiently high; hence the operative has the place where it forms well under control.

An incidental but important advantage of having a conductive base for the screen is to prevent the screen potential being affected should the tube be touched when, for example, measurements are being made. There is therefore no need for the screen to be at earth potential, as was usually the case in earlier types of oscilloscope tube. Indeed, in the new types the cathode of the electron gun can be earthed, so that the heater can be fed from one of the normal heater windings provided on the mains transformer.

Coatings of this kind generally give rise to interference colours. Their formation might be prevented by making the depth of the coating less than a quarter-wavelength of violet light, but so thin a coating would have insufficient conductivity for our purposes. Nor, on the other hand, can one make the coating so thick that interference would cease to play any part, because then it would not adhere to the glass. The coating that is in fact applied has a thickness such that no interference colours of an order higher than the first are produced. Furthermore, in order to avoid colour differences, this depth is kept as nearly constant as possible throughout the screen area. It is only the outward appearance of the tube that is concerned here; the interference effect makes no perceptible difference to the oscillogram itself.

The fluorescent powder must also be selected to suit the low anode voltage. From the brightness viewpoint the best phosphor is normal willemite (zinc orthosilicate activated with manganese in a Si/Mn ratio of 1 : 0.008). However, when a powder of normal composition is used, it is found that after about 600 hours the fall-off in brightness due to electron bombardment is steeper per unit of time than is considered desirable — see graph *a* in *fig. 1*. When an oscillogram on a tube of that age is examined, the points at which it intersects traces of earlier oscillograms are perceptible as points of lower brightness. The steep decline in brightness can be prevented by raising the manganese content; but then the initial brightness falls off. The relationship between the two quantities is shown graphically in *fig. 2*. The best compromise was found to be a

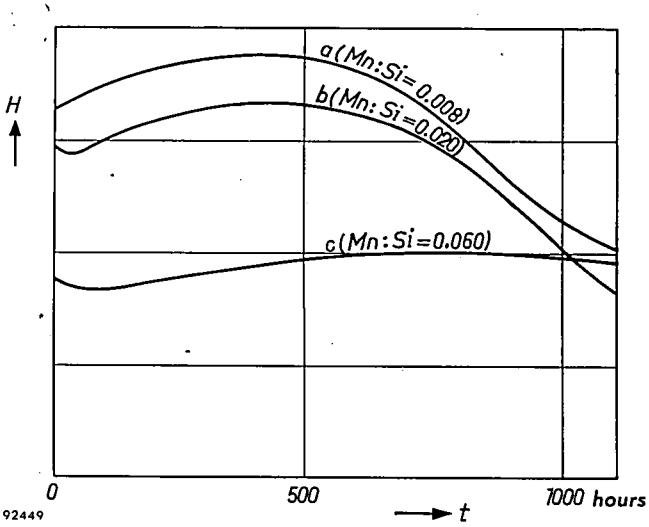


Fig. 1. Curves of H, the brightness of a trace on a willemite screen, as a function of time, when the screen is bombarded with electrons of 400 eV. The ratio of manganese to silicon atoms in the phosphor is 0.008 : 1 (the normal ratio) for curve *a*, 0.020 : 1 for curve *b* and 0.060 : 1 for curve *c*.

Mn/Si ratio of between 3 : 100 and 4 : 100. This gives curves lying in the region between curves *b* and *c* in fig. 1. The resulting degree of brightness is not unduly low (as it is in the case of curve *c*), nor does it fall off too quickly with time (as it does in curves *a* and *b*).

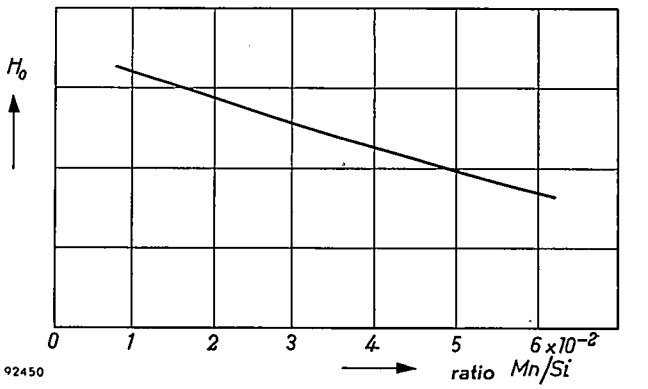


Fig. 2. Initial brightness H_0 of willemite screens as a function of the manganese content of the phosphor. The vertical axis is divided into arbitrary brightness units and the horizontal axis according to the ratio between manganese and silicon atoms. The curve shown is that obtained with a beam of 400 eV electrons.

The electrode assembly

A diagram showing the electrode assembly of the DG 7-32 appears in *fig. 3*. The assembly has three main parts, an electron gun delivering a beam of adjustable intensity, an electron lens that brings the image of the beam cross-over [4]) to a sharp focus on the screen, and a deflecting system whereby the beam is made to undergo the desired changes of direction.

[3]) See for example, H. te Gude and E. Schaaff, Philips tech. Rev. **18**, 243-245, 1956/57 (No. 8).

[4]) The cross-over is the point at which the diameter of the beam in the gun is smallest; it lies somewhere between the cathode and first anode. Refer to fig. 5 of article cited in footnote [5]).

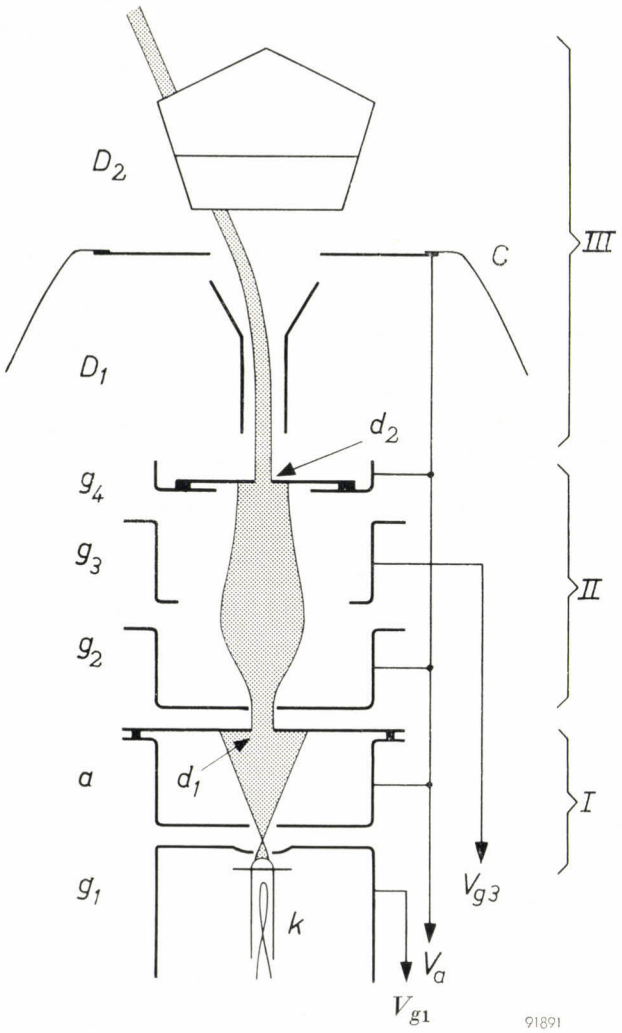


Fig. 3. Diagram to show the electrode assembly of a DG 7-32 oscilloscope tube. Three consecutive sections of the assembly may be distinguished, (*I*) the electron gun, (*II*) the electron lens, and (*III*) the deflecting system. Diaphragms d_1 and d_2 are electrically connected to the electrodes on which they are mounted. It will be observed that electrodes a, g_2 and g_3 have the same basic shape, a certain degree of standardization having been achieved in the parts making up the assembly.

A photograph of the tube, together with a complete and finished electrode assembly and the individual parts of which it is made up, appears in *fig. 4*.

Electron gun and lens

A triode gun is used, consisting of an indirectly heated cathode k, a control grid g_1 and a first anode a. The cathode is earthed, g_1 carries negative potential and the anode a has a potential of +400 V. The intensity of the beam can be regulated [5]) by changing the potential of g_1.

Electrodes g_2, g_3 and g_4 form the electron lens. By changing the potential of g_3 (the focussing electrode) the strength of the lens can be adjusted so as to give the sharpest possible image of the cross-over on the screen, that is, so as to produce the

[5]) Electron guns of all types are dealt with in J. C. Francken, J. de Gier and W. F. Nienhuis, Philips tech. Rev. **18**, 73-81, 1956/57 (No. 3).

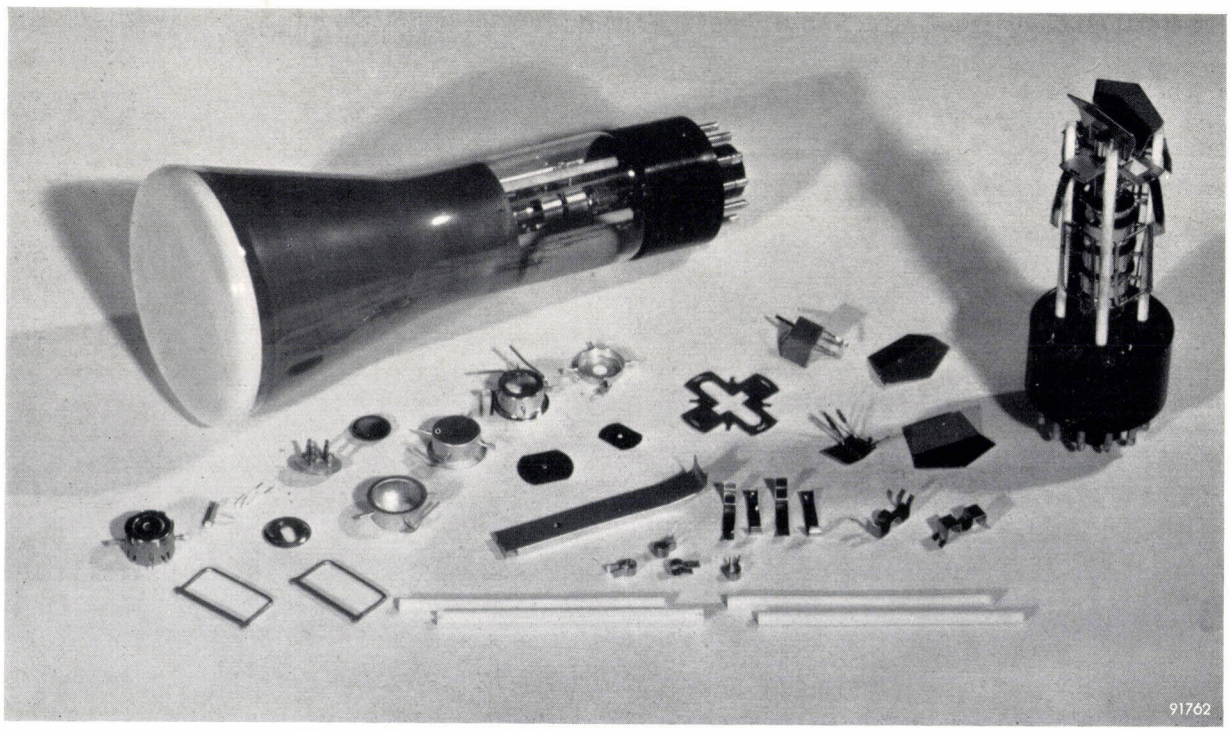


Fig. 4. A DG 7-32 tube, a finished electrode assembly and the individual parts comprising the assembly. The tangs on the grids and the D_1 plates can clearly be seen; these tangs are lodged in the powdered-glass filling of grooves in the ceramic rods.

smallest possible spot. The electrodes g_2 and g_4 have the same potential as a. The electrode g_3 carries a potential between 0 and +200 V, depending on the anode voltage.

It will be seen that in these tubes the functions of shaping and focussing the beam are distributed amongst several electrodes; this system compared with others that were formerly much used (systems with accelerating lenses), has the great advantage that the intensity and sharpness of the spot can be adjusted independently of each other.

Diaphragms d_1 and d_2 have the following function. d_1 constricts the beam issuing from the gun, thus preventing electrons striking the focussing grid g_3. Thus focussing is effected without a flow of current, making it permissible for the voltage source supplying g_3 to have a high internal resistance. Electrons which are scattered around the edges of d_1 and which might otherwise strike g_3 are intercepted by g_2. Diaphragm d_2 constricts the beam prior to its entry into the deflecting system.

Having been narrowed down thus by the agency of d_1 and d_2, the beam has only a small fraction of its original intensity on arrival at the screen. The loss of beam strength is particularly undesirable in a CRT with a low anode voltage, where the production of a sufficiently bright spot is a problem in itself. In order to reduce the loss the beam must be made as narrow as possible before it reaches d_1, and this may be done by strengthening the lens formed by g_1 and a. Doing so, however, is only too likely to lead to enlargement of the spot on the screen [6]). Enlargement of the spot may be prevented by bringing the lens nearer to the screen, but then the deflecting system has to be moved up too and this in its turn leads to a loss of deflection sensitivity. A compromise has therefore to be sought. In the tubes under discussion, the g_1-g_2 lens is given somewhat greater strength than the corresponding lens of the guns described in article [5]) by making a depression in the face of g_1.

Deflection system

The deflection system does not differ in principle from that which has been employed in oscilloscope tubes for some time now. We shall briefly discuss two details which have not received particular attention in previous articles. The two pairs of plates, pair D_1 for vertical deflection by the test voltage and pair D_2 for horizontal deflection, may be seen in fig. 3. In a tube with flat parallel deflecting plates (as in *fig. 5*) the formula for R, the displacement of the spot across the screen, is:

$$R = \frac{Ll}{2V_0 S} V_d, \quad \ldots\ldots \quad (1)$$

where V_0 is the difference of potential through which the electrons have fallen, V_d is the difference of potential between the plates, l the length of the plates, S their distance apart and L the distance from their centre to the screen.

It follows directly from the formula that, where V_0 and L and the screen diameter are given, the plate voltage required for maximum deflection can be lowered by increasing the ratio l/S. This can only be done to a limited extent, however, because a) the beam diameter d (see fig. 5) cannot be narrowed too much without reducing the brightness

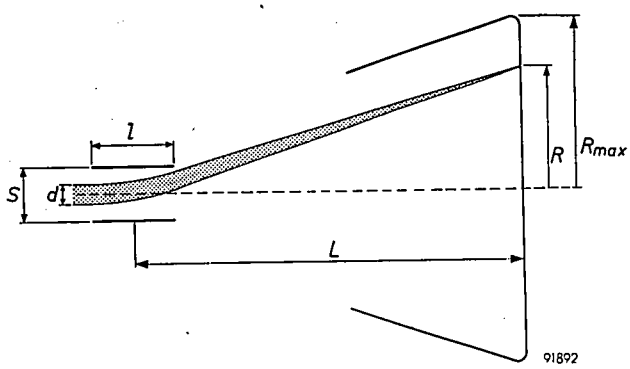


Fig. 5. The deflection of an electron beam by flat parallel plates.

of the spot, and b) the beam will soon start to be pulled against the plates if l is lengthened unduly, and this will result in distortion of the trace. The aim of lowering the deflecting voltage can nevertheless be achieved by bending the plates outwards; the bend in the D_1 plates can be seen in fig. 3, that in the D_2 plates in fig. 4.

An improvement of more recent date is a change in the shape of the D_2 plates; these are not rectangular, as before, but have the outline indicated in fig. 3 (see also fig. 4). The aim achieved by giving the D_2 plates this shape is that the path of the electrons between them has the same effective length whatever the degree of deflection they have previously undergone between the D_1 plates; in other words, the amount of deflection undergone in D_2 is independent of that undergone in D_1. Pincushion distortion of the image is thereby obviated.

One inference from formula (1) that is of particular importance for tubes with a low V_0 is that R/V_d, the sensitivity of the deflecting system, is inversely proportional to V_0 [7]). For previous types of CRT the minimum value of V_0 has been 800 V; the reduction of V_0 to 400 V, as in the present tubes,

[6]) A fuller account of "pre-focussing" may be found in the article cited in footnote [5]).

[7]) It is true that the formula has to be somewhat modified in respect of the bending of the plates, but this does not affect the part played by V_0.

therefore results in doubled deflection sensitivity. The average sensitivity of these tubes, both the DG 7-31 and the DG 7-32, (with V_0 at 400 V) is approx. 0.50 mm/V for the D_1 plates, which carry the test voltage, and approx. 0.30 mm/V for the D_2 plates.

Mounting of the electrodes

The electrodes are mounted on ceramic rods having a groove filled with powdered glass, in a manner that has already been described in this Review [9]). The rods are heated to a temperature such that the glass melts, permitting the tangs of the electrodes to be pressed into the grooves.

Besides acting as a shield, plate C (see fig. 3) serves for centering the electrode assembly and is provided with springs for that purpose. In the

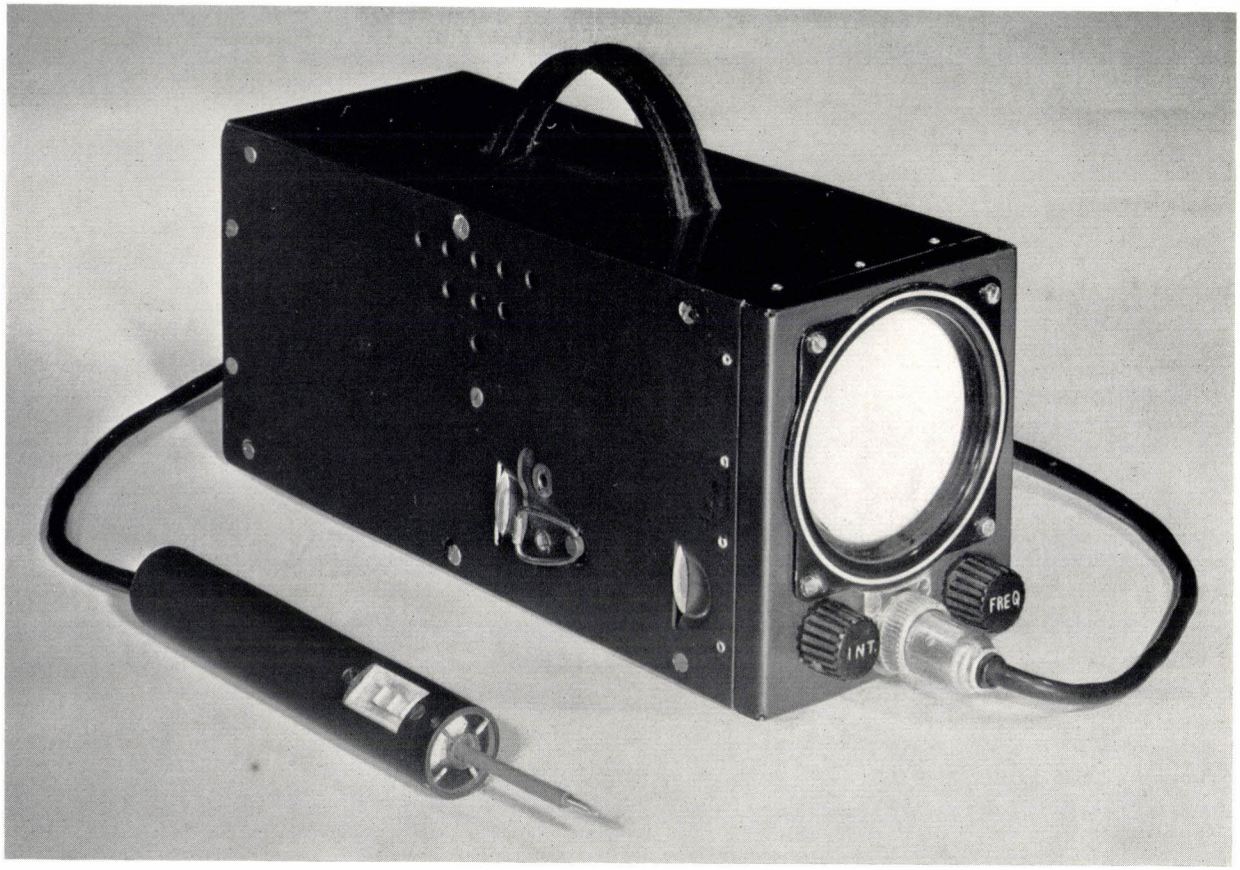


Fig. 6. An oscilloscope of small size, equipped with the DG 7-32 tube (this instrument is not on the market). The properties referred to in the text have been obtained for the oscilloscope in the following way. The incoming signal is passed through a two-stage amplifier equipped with two ECF 80 triode-pentodes. The triode portion of the first tube, connected as a cathode-follower, precedes the pentode portion; the second triode serves as a phase inverter. Even better properties can be conferred on the oscilloscope by using an attenuating probe in conjunction with it. In the first place, of course, the probe allows a stronger test signal to be admitted, but it also makes it possible to reduce the input capacitance to 5 pF. For this purpose the ECF 80 is replaced by a combination of an EF 80 pentode and an EC 92 triode. The latter tube, again as a cathode-follower, is embodied in the probe. With the CRT operating with the minimum sensitivity of 0.27 mm/V (i.e. with 400 V on the final anode), the amplitude of the sweep voltage required for a time axis 70 mm long is 126 V. The sweep voltage is generated by an EF 80 tube in a Miller transitron circuit [10]), one half of the ECC 81 double triode acting as phase inverter. The other triode acts as limiter for the synchronizing signal.

In the DG 7-31 tube, which has been designed for asymmetrical application of the sweep voltage, the unearthed plate has two metal rods fixed to it on the side nearest the D_1 plates. The function of these rods is to correct trapezium distortion due to the fact that the mean potential difference between the two pairs of plates D_1 and D_2 is not equal to zero [8]).

Asymmetrical application of the sweep voltage enables the circuit supplying that voltage to be simplified. It does, however, have the drawback that some slight distortion is unavoidable.

[8]) See J. D. Veegens, Philips tech. Rev. **4**, 198-204, 1939.

[9]) J. de Gier and A. P. van Rooy, Philips tech. Rev. **9**, 180-184, 1947/48.

[10]) See for example J. Czech, The cathode-ray oscilloscope, Philips Technical Library 1957, page 77.

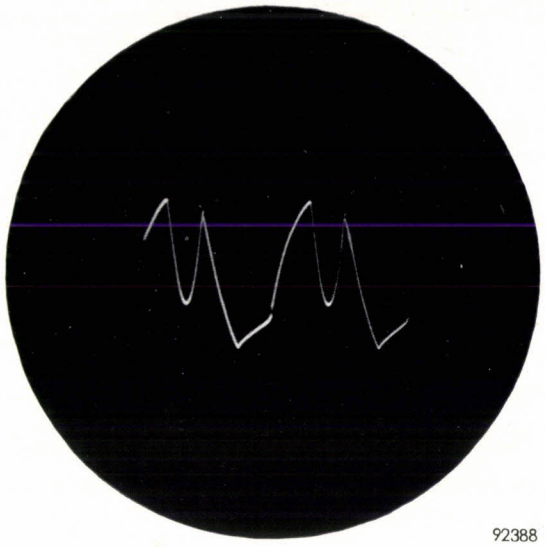


Fig. 7. An oscillogram displayed on one of the tubes described in the article (actual size).

DG 7-31 tube the compensating rods fixed to D_2 prevent plate C being mounted in the place indicated in fig. 3. In this tube, therefore, the centering plate is combined with grid g_4. Apart from this, the electrode assemblies of the two tubes are the same.

As already stated, the tubes just described can be made the basis of a very compact and lightweight oscilloscope such as may serve the purposes of television servicing, for example. A photograph of a small instrument of this kind, equipped with the DG 7-32 tube, appears in *fig. 6*. Its dimensions are only $10 \times 12 \times 27$ cm, it weighs 2.6 kg and its power consumption is as low as 26 watts [11]).

With a test voltage amplifier having a bandwidth [12]) of 3 Mc/s, the oscilloscope has a sensitivity of 100 mV per cm. The sweep frequency can be varied from 20 c/s to 25 kc/s. The maximum amplitude permissible in the input signal is 15 V. Input capacitance is 20 pF.

An oscillogram has been reproduced in *fig. 7* to give an impression of the performance of these tubes. In particular, the extremely good definition of the trace is noteworthy.

[11]) The instrument was developed by J. Stolk and W. D. Minjon in the Eindhoven applications laboratory for cathode-ray tubes.

[12]) Defined, as is usual, as the width of the band between those frequencies at which the amplitude falls off by 3 dB.

Summary. An account is given of the development of oscilloscope tubes requiring an anode voltage of only 400 V. There are two types, DG 7-32 and DG 7-31. The sweep voltage is fed to the first type symmetrically; one X-plate is earthed in the second type, the sweep voltage therefore being fed asymmetrically. The fluorescent screen is applied by the flow-coating method, and not with the aid of an inorganic binder. Such a layer of binder, covering the screen, would decelerate the beam electrons excessively. To prevent the screen from becoming charged, the screen portion of the envelope is first given a conductive coating of tin oxide. The coating offers two additional advantages: in the first place, when measurements are being made, the glass can be touched without the screen potential being affected; secondly, it is possible for the cathode of the tube to be earthed, so that its heater can be fed from one of the normal heater windings on the mains transformer. The electrode assembly has been designed in such a way that focussing of the beam and adjustment of its intensity can be carried out independently of each other. The spot produced on the screen is small, and the definition of the trace therefore good. Two improvements to the deflecting system, introduced some time ago, are dealt with. Finally, some details are given of an oscilloscope of very light weight and of very small dimensions, the construction of which has been rendered possible by the availability of the tubes here described. The possibility of making such an instrument will be of particular interest to the radio and television serviceman, and also to the amateur equipment builder.

THE FORMULATION OF PESTICIDES

by W. DUYFJES *).

632.95

Agriculture and horticulture are increasingly exposed to plagues of such organisms as insects, mites, fungi and weeds. The chemical means of combating such pests come under the general heading of "pesticides". In previous volumes of this journal, articles have described the research work conducted by N.V. Philips-Roxane in the quest for new pesticides. The value of a pesticide depends not only on the active substance; the "formulation" of the preparation is equally important. By formulation is understood the making up of the active substance in a form (solution, emulsion, suspension, etc.) in which it is most efficacious, stable in transport and storage, reasonable in price, and so on. The problems relating to formulation are the subject of the present article and are studied in the formulation laboratory at the Amsterdam factory of N.V. Philips-Roxane.

The chemical control of agricultural pests

The great increase in world population in recent decades is largely to be attributed to the rapid advance of science in the widest sense of the word. Chemistry has had a large share in that advance: new medicaments have saved countless human lives, and the use of artificial fertilizers has made it possible to put vast barren regions under cultivation. This intensification of agriculture and horticulture has lead to considerably increased food production, but unfortunately it has also had an undesirable consequence: injurious organisms, such as certain insects, red spiders, moulds and weeds have found new breeding grounds and threaten to rob man of what he has newly acquired. Although small mammals, birds, amphibians, reptiles and spiders take enormous toll of injurious insects, this is not nearly enough to avert the danger. Mechanical methods of extermination were soon found to be no longer an economic proposition, and thus man has fallen back on chemical agents: pesticides. This collective name covers insecticides, acaricides, fungicides and herbicides, specific as their names imply, for a certain type of pest. Chemicals were used in the last century on a very modest scale for destroying insects, moulds and weeds. But it was the discovery by Müller in 1939 of the insecticidal activity of DDT (dichloro-diphenyl-trichloroethane) that gave fillip to the large-scale comparative investigations into the biological activity of known compounds. New insecticides discovered in this search have largely superseded the "natural" insecticides (nicotine, rotenone, pyrethrins) as well as the synthetic products, such as inorganic arsenates and fluorides used prior to 1939, and have given birth to a new chemical industry. The latter supplies thousands of tons of pesticides a year which contribute towards safeguarding the harvest.

As regards the control of insects, it was found that the presence of chlorine in an organic molecule could lead in certain cases to insecticidal activity. As a corollary to this observation, such substances as BHC (benzene hexachloride), heptachlor, toxaphene, aldrin, dieldrin, endrin, etc. came into use, following upon the discovery of DDT.

A new group of insecticides that was discovered comprised esters of phosphoric and thiophosphoric acids. These form the basis of such substances as tetraethyl pyrophosphate, parathion, malathion, etc., which have rapidly gained ground. The surprising activity, as it was initially described, of DDT and other chlorine-containing insecticides is even surpassed by that of the phosphates and thiophosphates, since they have a wider action spectrum, being also effective against red spider. This was discovered by Schrader just in time, for the red spider had multiplied enormously owing to the use of chlorinated hydrocarbons (DDT, BHC, lindane, etc.) which had decimated its natural enemies.

Such upsets of biological equilibrium resulting from extermination of the enemies of noxious animals, together with the resistance phenomena that have been observed, have led to the realization that the frequently rigorous methods used for combating insects make an early solution to the insect problem unlikely.

*) N.V. Philips-Roxane, Pharmaceutisch-chemische Industrie "Duphar", Amsterdam.

If, initially, the search was for substances that would kill as many types of injurious organisms as possible, the experience gained soon showed that substances with a more selective action, and above all a more intelligent employment of agro-biologicals, made it more probable that the desired goal would be reached. Thus, for example, emphytic (sometimes called systemic) agents have been developed, such as "Systox" and "Pestox", which are assimilated by the plant, rendering it toxic to the parasite. Further, specific ovicides have been discovered: e.g. "Tedion V18" (2,4,5,4′-tetrachlorodiphenyl sulphone) [1]) for exterminating red spider eggs; and also insecticides and acaricides with other active groups in the molecule. These new substances were developed not only with a view to getting round certain disadvantages of the known substances, e.g. their tendency to arouse resistance, their toxicity to man and domestic animals, but mainly with the intention of sparing useful organisms (birds, spiders and other natural enemies of the pest), so as to disturb the complex balance of nature as little as possible.

Good results have also been recorded in the combating of fungi and moulds [2]). Fungicides, such as sulphur, sulphur compounds and copper compounds which before 1939 were the only ones in use, have been supplemented by another important group comprising the dithiocarbaminates, organic mercury compounds, trichloromethyl derivatives and triazines. Fungicides such as the proprietory products zineb and captan have made general headway. Thus far there have been no indications that the new fungicides will create problems like those arising with the insecticides.

The same applies to weed control. Rapid progress in this field is the outcome of the discovery of the "growth-substance herbicides" [3]), which are marketed under names such as MCPA, 2,4-D and 2,4,5-T.

The importance of the right "formulation"

In practice a relatively small amount of the active substance, the pesticide, must be distributed over the plants that are to be protected against animal pests or moulds, or over the weeds that are to be destroyed. The pesticide can be distributed over a large area by sprinkling or spraying (a liquid), by dusting (a dry powder) or by atomizing (a liquid). Accordingly, we thus speak of sprays, dusts and aerosols.

In the preparation of a product for one of these modes of application, the pesticide must be mixed with various accessory substances to give a proprietary preparation which is either ready for use or can be made up for use in a simple manner (e.g. by adding water). The main accessory substance is the carrier; others, as we shall see, are wetting agents, dispersing agents, and stickers.

The selection of these substances and the compounding of them, i.e. the formulation of the active ingredient with certain other substances to give a preparation combining the best properties, is of the utmost importance. As a first requirement, an agro-biological preparation must be capable of being distributed over a large area in one of the above-mentioned ways. Moreover it must not corrode, block or cause wear to the nozzle of the equipment. Once on the plant, the preparation must adhere long enough to do its work; it must not be too volatile therefore; nor must it drain off too quickly or be removed by the wind and rain. Naturally the preparation must not deteriorate in storage and should be relatively inexpensive.

Very briefly we shall now describe the train of events in pesticide research at Philips-Roxane [4]). In the Weesp (Holland) laboratory many substances are synthesized which might conceivably possess a pesticidal action. These substances are studied in the "Boekesteyn" agro-biological laboratory for their pesticidal action and those that show promise as insecticides or fungicides are tested for their phytotoxicity. Their toxicity to warm-blooded animals is likewise ascertained. If the results of these tests are favourable (and this is only very seldom the case!) further biological tests and preliminary field assays are undertaken with the substance compounded in a provisional formulation. If the results are satisfactory here also, and it is decided to manufacture the compounds, the question of the proper formulation arises. The form of the preparation and the choice of accessory components can only be arrived at when various properties of the active ingredient are known: vapour pressure, melting point, solubility, milling properties, sensitivity to light, tendency to reduction or oxidation, crystallization properties, hydrolyzability, etc.

Incorrect formulation will make a useless product of a good pesticide. The composition of a prepara-

[1]) J. Meltzer, Research on the control of animal pests, Philips tech. Rev. **17**, 146-152, 1955/56.

[2]) M. J. Koopmans, Fungicide research, Philips tech. Rev. **17**, 222-229, 1955/56.

[3]) R. van der Veen, Growth substances in plants, Philips tech. Rev. **17**, 294-298, 1955/56.

[4]) R. van der Veen, "Boekesteyn", the agro-biological laboratory of N.V. Philips-Roxane, Philips tech. Rev. **16**, 353-359, 1954/55, in particular pp. 357-358.

tion therefore calls for special attention. A survey of the principal problems that arise in formulation now follows.

General considerations on formulation

Particle size

A pesticide can only answer its purpose if the organism it is to control comes into contact with the active substance and takes up sufficient of it. Both conditions are best met if the substance is finely divided. The smaller the particles, the greater their active surface. For example, if grains with an initial diameter of 1 mm are ground down to particles having a volume 10^6 times smaller, i.e. a diameter of 10 μ, these new particles will have a collective surface 100 times greater than the collective surface of the original grains. The probability that the compound will come into contact with the organism is increased to the same extent. Moreover the organism can more readily take up the smaller particles.

How small the particles must be in order to be sufficiently active depends upon all kinds of factors: the active substance, the organism it is to control, the manner in which it is to be applied and the climatic conditions. In general, however, it can be said that preperations in which the particles of active substance are greater than 50 μ, will not be very effective. This does not of course apply where the preparation is to be used in the form of a (true) solution. 10 to 30 μ is usually a satisfactory diameter for insecticide particles: because of their usually high mobility, insects take up a relatively large amount of a given substance. For the control of mites and particularly their larvae or eggs a finer division must be aimed at, since these organisms are far smaller than most insects and less mobile.

The combating of fungi calls for still greater dispersion. A particle size of only a few μ is the norm for fungicides that are to be applied in suspension or as a dust.

A very fine state of division of the substance has disadvantages, however. The increase in surface that attends finer division makes for more rapid evaporation. This is especially important in the case of insecticides having a residual activity and which must leave a residue on the plant. Some of these have an appreciable vapour pressure, thus too fine a division of the preparation would mean that the protective action was of too short duration. (It should be remembered in this connection that milling never gives a uniform product: if a product is ground to a mean particle size of say 10 μ, a large number of much smaller particles will be formed which evaporate relatively quickly.)

Volatility can, moreover, bring numerous undesirable reactions to light. If there is any "chemical incompatibility" between the accessory components (more about this later) or some degree of toxicity to the plant which is to be protected, these effects will be magnified with smaller particles. This is true also of the lumping of powder preparations when stored for lengthy periods, especially under heightened pressure (stacked up in bags) and tropical temperatures. The lumping can lead to blockage of the nozzle; another ingredient is therefore necessary to prevent lumping. The World Health Organization has laid down standards to which certain agro-biological preparations have to conform. They include a "tropical storage test"; the success with which a preparation survives this test gives an idea of its ability to stand up to transport and storage in the tropics.

The carrier

The carrier has already been mentioned as the most important accessory component. Its function in many cases is to afford better distribution of the active ingredient. Frequently it also serves as the grinding medium and as an adsorbant which gradually dispenses the active ingredient (this process must be neither too quick nor too slow). Laboratory tests can give valuable indications as regards the rate of dispensation, but in the case of powder preparations it is always necessary to make biological tests to ascertain whether the activity is near enough to the optimum.

In the development of a preparation, the "chemical incompatibility" alluded to above, must be taken into account. An insecticide such as DDT can react with OH-ions, splitting off HCl to give a compound which is biologically inactive. Thus such an insecticide must not in general be formulated with an alkaline carrier (in certain cases however basic carriers are nevertheless compatible with pesticides that split off HCl). Conversely a pesticide may undergo chemical decomposition owing to certain impurities in the carrier which initiate catalytic decomposition. For the above reasons, a compatibility test — which is simple and rapid to carry out — should never be omitted.

Some carriers contain abrasive ingredients. If they possess a certain structure and grain size, they can do such damage to the insect cuticula that water is lost by the insect through evaporation,

with resultant death of the insect. Such carriers therefore possess an insecticidal action in themselves, but are nevertheless highly undesirable, killing useful insects and causing mechanical damage to the plant, and rapid wear of the dusting equipment and spray nozzles.

Powder preparations can be divided into dusts and wettable powders. These differ widely in their contents of active ingredient: dusts are preparations ready for use and thus possess a fairly low content of active substance; on the other hand wettable powders are suspended in water before use and the actual ("secondary") carrier is thus water. For reasons of economy (freight costs), the content of active substance in wettable powders is made as high as possible. The ("primary") carrier present in the powder must thus possess a much greater adsorbing capacity, and consequently the danger of lumping (see above) is appreciably more imminent. The formulation of wettable powders therefore demands greater care than the formulation of dusts.

Other ingredients

On account of the water-repellent nature of many pesticides their suspension in water calls for a wetting agent, i.e. a substance that lowers the surface tension of water. This wetting agent must also be compatible with the active substance and with both the primary and the secondary carriers (the latter is water). Soap — the classical surface-active agent — is unsuitable because of its susceptibility to Ca and Mg ions. Saponin and similar natural substances have been completely ousted in the agrobiological industry by the cheaper synthetic compounds, such as fatty alcohol sulphates, alkylaryl sulphonates, polyethylene glycols, etc. One or two per cent of such substances is sufficient to render a preparation wettable.

The wetting-agent content of wettable powders must be such as to ensure that the liquid spray prepared from the powder and water retains sufficient surface tension to prevent rapid drainage of the pesticide from the plant. Nor should the surface tension be so low that troublesome foaming occurs in the spray tank. Further, certain surface-active agents enhance the efficacy of the active substances, so that such agents should be chosen not only for their physico-chemical but also for their biological properties.

In addition to a wetting agent, a wettable powder usually requires a dispersing agent to counter flocculation, i.e. the aggregation of the finely divided particles. Flocculation would lead to rapid sedimentation of the suspended material and this would be attended with a reduction in the active surface and with blockage of the spray nozzle.

In certain cases wettable powders must also contain a sticker. The latter prolongs the action time, this being especially useful in the case of poorly adhering fungicides and insecticides that work as stomach poisons (see the article cited in [1]), page 146). A highly tenacious residue is undesirable however if the active compound is very stable and particularly toxic to man and domestic animals.

Classification of pesticide preparations

Above we have made a distinction between sprays, dusts and aerosols, according to the way in which a pesticide preparation is applied to the plants. The first and the third groups can be further subdivided, whereupon we arrive at the following classification:

- Sprays
  - solutions
  - emulsions
  - solubilized solutions
  - suspensions
- Dust preparations
- Aerosols
  - smoke generators
  - liquid preparations for atomization
  - fumigants

We shall now deal with the principles whereby these different groups are formulated.

Sprays

Solutions

A "true" solution (in contrast to an emulsion or a suspension) gives the best distribution of a pesticide. The insects, mites, fungi or plants to be exterminated are soaked with the pesticide, and when the solvent has evaporated, are left with a homogeneous layer or "crystal carpet" of the pesticide.

Spraying agents intended for use in solution are on the market in both liquid and powder form. The former need only be made up to the right concentration with water (in a few rare instances with paraffin oil) before use; the powder products have to be dissolved in water. To prevent undesirable reactions with hard water, the preparation frequently contains certain other additives. Beyond this, the formulation of water-soluble pesticides presents no serious problems.

Modern pesticides for the control or insects or mites, such as DDT, BHC, lindane, "Tedion V18" and many others are but slightly soluble in water

621.375.2.029:6:621.385.3.029.6

J. P. M. Gieles: A 4000 Mc/s wide-band amplifier using a disc-seal triode,
Philips tech. Rev. **19**, 145-156, 1957/58 (No. 5).

Description of an amplifier for frequencies from 3800 to 4200 Mc/s, in which tubes EC 56 or EC 57 can be used. With a gain of 8 dB, the amplifier can deliver a power of 0.5 W and 1.5 W respectively, the bandwidth being 55 Mc/s. The low-level gain is 12 and 13 dB respectively. The input and output can be connected to waveguides. For coupling several of these amplifiers in cascade, the coupling is preferably effected by means of ferrite isolators. The variation of group delay in the frequency band covered is much smaller than with the IF amplifiers normally used in beam transmitters.

621.397.62

J. J. Balder: A luminous frame around the television screen,
Philips tech. Rev. **19**, 156-158, 1957/58 (No. 5).

Observers viewing a picture screen were asked to adjust the brightness and width of a uniformly illuminated frame around the picture so as to give most agreeable viewing conditions. This experiment was made with 20-25 observers each under conditions of a number of luminance values of screen and surroundings. The preferred frame width was found to be independent of screen and ambient luminance and averaged about 0.3 times the half screen dimensions for all observers. The preferred frame luminance was found to increase with the screen luminance as well as with the luminance of the surroundings. In order to come closest to the majority preference, it appears that the frame luminance should be adjustable between 0 and 40 cd/m^2.

621.317.755

F. de Boer and W. F. Nienhuis: Low-voltage oscilloscope tubes,
Philips tech. Rev. **19**, 159-164, 1957/58 (No. 5).

An account is given of the development of two oscilloscope tubes (types DG 7-32 and DG 7-31) requiring an anode voltage of only 400 V. One tube has been designed for symmetrical, the other for asymmetrical application of the sweep voltage. The fluorescent layer is applied on top of a conductive tin oxide coating (to prevent the screen charging up) and is not covered by a binder layer. Such a layer would decelerate the beam electrons excessively. Thanks to the conductive coating, it is possible to touch the tube while measurements are being made; furthermore, the cathode can be earthed, so that the heater can be fed from a normal heater winding in the transformer. The electrode system has been designed in such a way as to make the adjustment of brightness and focus independent of each other. Definition is very good. The tubes lend themselves well to the building of small lightweight oscilloscopes, and one such instrument is discussed in the article.

632.95

W. Duyfjes: The formulation of pesticides,
Philips tech. Rev. **19**, 165-176, 1957/58 (No. 5).

Animal and vegetable pests are combated chemically using pesticides (insecticides, acaricides, fungicides, herbicides). An agro-biological preparation consists of a substance having a pesticidal action and various auxiliary substances, including a carrier. The formulation of these preparations is of the utmost importance. According to the way in which they are used, preparations are classified as sprays, dusts and aerosols. In this article the main questions that arise in the formulation of such pesticides are discussed.

and cannot be used in the form of true solutions. If these compounds are to be sprayed with water, procedures which will be described in the section on emulsions below, must be adopted. We may mention nicotine (now little used) as one of the few insecticides sufficiently soluble in water for normal application.

Because of its phenolic structure, DNC (4,6-dinitro-orthocresol) which has been used since 1892, occupies a special place. DNC is an acid with a low water solubility, but salts which dissolve satisfactorily in water can be prepared by replacing the "acid" H-atom by Na, K or NH_4. These salts are biologically less active than the acid; the ammonium salt is the most effective, since after spraying it decomposes fairly quickly into the more active DNC-acid and NH_3. DNC and its salts are both insecticidal (ovicidal) and herbicidal. But they are also phytotoxic and thus can be employed to kill aphid eggs on fruit trees only in the winter, i.e. during the dormant period of the trees.

We may mention lime sulphur spray (calcium polysulphide) and certain organic mercury compounds as examples of the few fungicides which are sufficiently soluble in water. Well-known fungicides like copper oxychloride, captan, zineb and many others are sprayed in suspension because of their poor solubility in water (see section on suspensions).

A greater number of herbicides are sufficiently soluble in water. Here we shall confine ourselves to the growth substance herbicides [3]). Most of these are derivatives of phenoxyacetic acid. The principal derivatives have already been referred to: MCPA (2-methyl, 4-chlorophenoxyacetic acid), 2,4-D (2,4-dichlorophenoxyacetic acid) and 2,4,5-T (2,4,5-trichlorophenoxyacetic acid). They exhibit remarkable selectivity between mono- and dicotyledons, and because of this can be used to destroy dicotyledonous weeds growing among monocotyledonous cereals or grasses, without doing any appreciable harm to the latter. To be sure of complete success certain precautions must be taken in their formulation as will be evident from the following.

MCPA preparations are mostly marketed as a concentrated solution of the Na, K or NH_4 salt in water. Since the salts of bivalent metals are sparingly soluble in water, Ca- and Mg-methylchlorophenoxyacetate may be precipitated if the concentrated solution is diluted with hard water. This can lead to blockage of the sprayer and, furthermore, these acetates have no herbicidal action, so that part of the MCPA is inactivated. It is therefore desirable to add sequestering agents during formulation, i.e. substances which form a complex with Ca and Mg ions so that they can no longer react with the phenoxyacetate. Synthetic dispersing agents are also used for this purpose. But it is important not to add too much of such agents, otherwise the spraying liquid will wet the plants too thoroughly, whereupon monocotyledons may also suffer injury.

The alkali salts of 2,4-D are relatively poorly soluble in water and are thus marketed in powder form only, being sprayed with water in a concentration of a few tenths of a percent. However, the amine salts of 2,4-D are very readily soluble in water; a 40% solution of 2,4-D in the form of the triethanolamine salt is very stable and will withstand a moderate frost. The mono- and di-ethanolamine salts can also be used, also the mono-, di- and trimethylamine and morpholine salts. The correct *p*H value is important with these concentrated salt solutions. Such preparations are very easy to make up.

The alkali salts of 2,4,5-T are still less soluble in water than those of 2,4-D and for this reason are not used. The organic amine salts of 2,4,5-T are reasonably soluble, but have not been applied on any large scale, mainly because certain 2,4,5-T esters [5]) have an appreciably better herbicidal action. Since these esters are not sufficiently soluble in water, they must be formulated in a special manner, which will be discussed in the following section.

Emulsions

The insecticidal action of mineral oils has been known for some 200 years, but is was less than a century ago that stable emulsions of petroleum hydrocarbons in water were successfuly prepared. Such emulsions are two-phase systems consisting of water in which the oil is dispersed as fine droplets. It was about 80 years ago that the first emulsified preparations came on to the market for use against certain scale insects and red spider.

The shortcomings of these products were many and mainly a consequence of the incompatibility of the emulsifier then used — soap — with hard water, which caused the emulsion to separate into oil and water ("breakdown"). When emulsifiers became available which did not suffer from these disadvantages, e.g. protein substances (such as casein), cellulose derivatives and especially the synthetic detergents, mineral oil emulsions could be prepared which have done excellent service to this day in the control of red spider, especially in citrus fruit farming.

It has likewise been found that if an emulsifier is dissolved in a mineral oil or in a tar oil distillation

[5]) Philips tech. Rev. **18**, 335, 1956/57 (No. 11).

product, the solution will spontaneously emulsify in water. This has been applied on a large scale in what are called winter washes for combating aphid eggs. Large quantities of tar oils are employed each year for that purpose, notwithstanding the advantages of DNC (see above).

There are two ways in which an emulsive spray can be made up by the user:

1) by diluting a concentrated stock emulsion with water,
2) by pouring an emulsifiable concentrate into water.

Stock emulsions. Concentrated emulsions of mineral oils in water are fairly easy to prepare. One common technique is to add the oil phase discontinuously under constant stirring, to the water phase in which an emulsifier (usually a colloid such as ammonium caseinate) has been dissolved, until a very viscous emulsion is obtained. The latter is then adjusted to the right concentration with water. In this way stock emulsions can be manufactured having a mineral oil content of 80% which are still fluid. The emulsifier content can be very low, e.g. 0.5%. The advantage of a low emulsifier content is that the emulsion soon breaks down after spraying; the active oil is left adhering to the crop, while the water drains or evaporates off.

If the oil phase of the emulsion contains another substance, such as, for example, DNC, care must be taken that the emulsion has a low *p*H value, to preclude migration of the DNC from the oil phase to the water phase. This readily occurs in emulsions stabilized with ammonia, and results in the formation of ammonium dinitrocresolate in the water phase. When the emulsion breaks down, part of the active substance is lost with the water that drains off, especially if the spray liquid has a low surface tension.

Another advantage of stock emulsions is that they can be diluted with either soft or hard water; this is a point in their favour over the emulsifiable concentrates. By introducing certain additives special effects can be achieved: prolonged action time, compatibility with fungicides, insecticides or acaricides of quite different formulation, etc.

Over and against these advantages there are certain drawbacks. In the first place we may mention their susceptibility to changes in temperature and their liability to freeze. However, this can be remedied by choosing an appropriate emulsifier; emulsions have been prepared capable of withstanding both frost and tropical temperatures. Another drawback is the danger of undesirable reactions with the container. A serious consideration also is that many valuable pesticides cannot be made up into stock emulsions because of their readiness to hydrolyze, e.g. a number of thiophosphoric acid esters (malathion, parathion), chlorinated hydrocarbons with labile Cl atoms (toxaphene, dieldrin, aldrin) and esters of 2,4-D and 2,4,5-T (see above).

Because of these disadvantages the stock emulsions have slowly been ousted by the emulsifiable concentrates.

In two cases, however, they have completely held their ground.

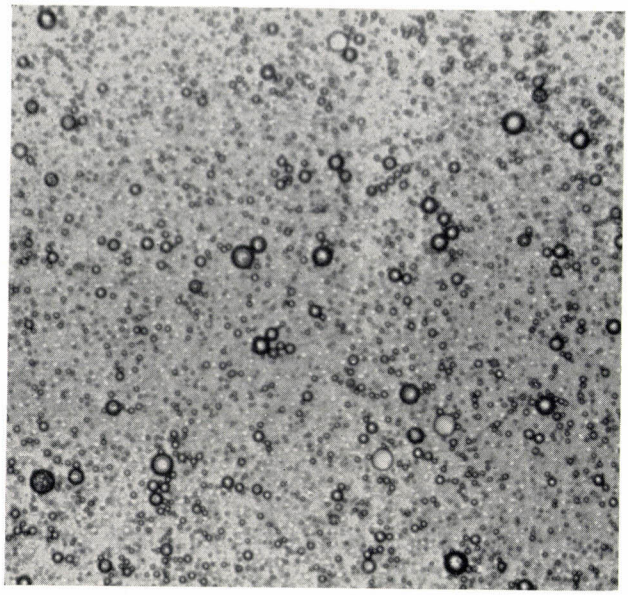
1) "White oil", an emulsion of refined mineral oil, is still very widely used, since its phytotoxicity is less than that of emulsifiable concentrates. Perhaps this can be attributed to an "alleviating action" on the part of the emulsifier: when the emulsion breaks down, the heavier water layer containing the colloid may perhaps prevent the oil from coming into direct contact with the plant.

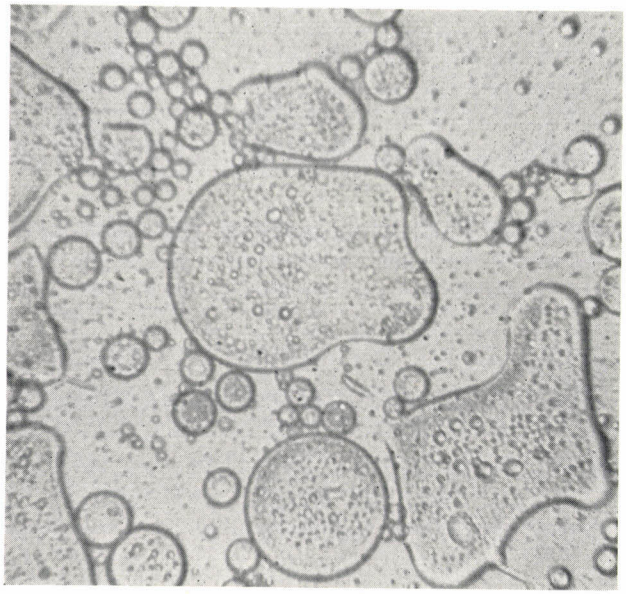
2) Many strong acidic pesticides are easier to formulate as stock emulsions than as emulsifiable concentrates, e.g. pentachlorophenol; one use of this substance is as a defoliant; it quickly rids potato plants of their foliage (this considerably simplifies lifting and likewise counters any infection with potato blight). Well-formulated pentachlorophenol emulsions rapidly break down after application, as is required for this purpose.

Emulsifiable concentrates. These are one-phase systems which spontaneously emulsify on mixing with water, the oil dispersing as droplets (discontinuous phase) throughout the water (continuous phase). The micro-photos in *fig. 1* show: the emulsion formed from an emulsifiable concentrate; the breaking-down of the emulsion; and the residue. In their simplest form, emulsifiable concentrates consist of two components: an active liquid and, dissolved in the latter, an emulsifier. As we shall see shortly, however, more complex compositions are required in practice.

The molecules of an emulsifier contain a hydrophobic part having an affinity for the oil phase, and a hydrophilic part having an affinity for the aqueous phase. When an emulsifiable concentrate is poured into water the emulsifier molecules orientate themselves; they form links at the oil-water interfaces between the two phases which were not originally miscible. The emulsion then forms at once, and is usually of a dual nature, i.e. microscopic examination reveals a dispersed aqueous phase within the dispersed oil phase (fig. 1); this is of no practical significance.

Modern emulsifiers employed in the preparation of emulsifiable concentrates almost always consist of two surface-active substances, one of which forms no ions while the other forms anions at the





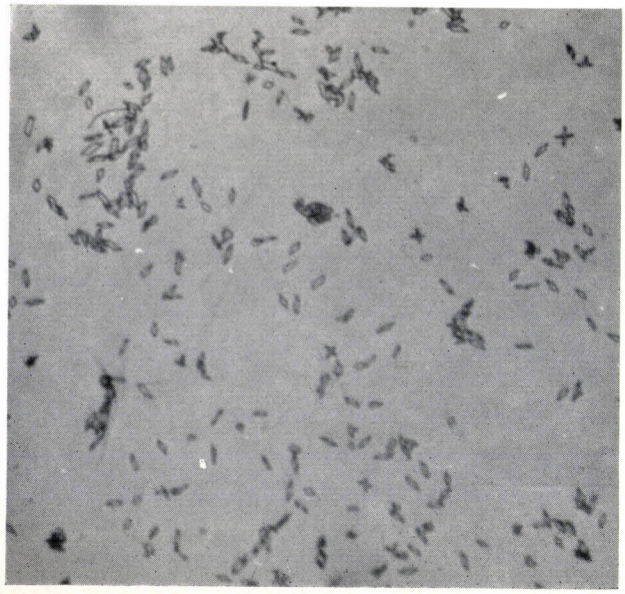


Fig. 1. Photo-micrographs of the acaricide "Tedion V18", formulated as emulsifiable concentrate. Magnification 500×. *Above:* Emulsion of the emulsifiable concentrate in water. *Centre:* The emulsion is breaking down. Water droplets are emulsified in the oil droplets (still surrounded by water), which are by now fairly large. *Below:* After the emulsion has broken down and the solvent evaporated, the residue is left in the form of fine crystals.

interface. These substances must be chosen individually for each case and must be accurately matched, so that the best possible results are obtained with water of varying hardness. Often a stabilizer is necessary to protect the emulsifier from reaction with decomposition products of the active substance (see below), thus ensuring that the emulsifiability is retained.

The emulsifier is chosen not only for its physical and chemical properties but also for its biological qualities. Some emulsifiers are rather toxic to the plant. When the emulsion has been sprayed, the oil phase that separates out will still contain a fraction of the emulsifier; through its affinity for hydrophobic and hydrophilic interfaces, this emulsifier will penetrate deeper into the plant than the active substance itself (which is usually hydrophobic) and can thus do quite a lot of damage to sensitive crops.

There is another danger to the plant besides the emulsifier; on decomposing, the active substance may give phytotoxic decomposition products. Decomposition can be caused by the presence of too much water or of metals (iron or tin containers). For this reason the water content of emulsifiable concentrates must be kept as low as possible (0.2% is permissible in favourable cases), and metal containers must have a good synthetic-resin lining.

Besides an emulsifier and a stabilizer, an emulsifiable concentrate may require yet another component, namely a solvent for the active substance if this is solid or highly viscous. The solvent is usually a mixture of two or more liquid hydrocarbons (xylene, paraffinic oils).

A widely practised technique, but one that is not to be recommended, is to spray a combination of agro-biological preparations, e.g. an insecticidal emulsion and a fungicidal suspension. The results of such treatment are frequently disappointing, especially if the suspension contains strongly hydrophobic active ingredients. For example, an emulsion of parathion combined with a suspension of zineb shows a strong tendency to flocculate. This soon leads to blockage of the nozzle and likewise reduces the activity of both substances. In general therefore the combinations of emulsions and suspensions should be discouraged.

It is possible to formulate emulsifiable concentrates so that they are physically compatible with other preparations (including suspensions). This entails the mixing of a large amount of strongly hydrophilic emulsifier in the emulsifiable concentrate and it is thus doubtful whether such preparations are an economic proposition.

Solubilized solutions

Solubilized solutions are clear solutions of certain substances in water, being one-phase systems composed of a substance immiscible with water, one or more surface-active agents and certain co-solvents, in the correct proportions. Preparations of this type are characterized by a high emulsifier content (usually predominantly hydrophilic); the co-solvents are usually alcohols. The interfacial tension between the active substance and water is then near enough zero, i.e. there is well-nigh perfect miscibility. The active ingredient in such preparations is particularly finely divided (particles varying from a few tenths to a few thousandths of a micron).

Solubilized DDT preparations have been employed, but have few advantages to offer. On the other hand, solubilized preparations of emphytic agents are important. Naturally both the active substance and the emulsifier must be perfectly innocuous to the plant itself; moreover any breakdown products toxic to man or domestic animals must not accumulate in parts of the plant that are destined for consumption. Here, therefore, there must be a far-reaching investigation into plant and animal toxicity.

The use of the growth substance herbicides in the solubilized state has made little headway. It is to be expected that esters of 2,4-D or 2,4,5-T in the solubilized form will be assimilated by weeds more rapidly than the residue from the usual rapidly breaking-down emulsions. Further research will have to show which formulation is to be preferred.

Suspensions

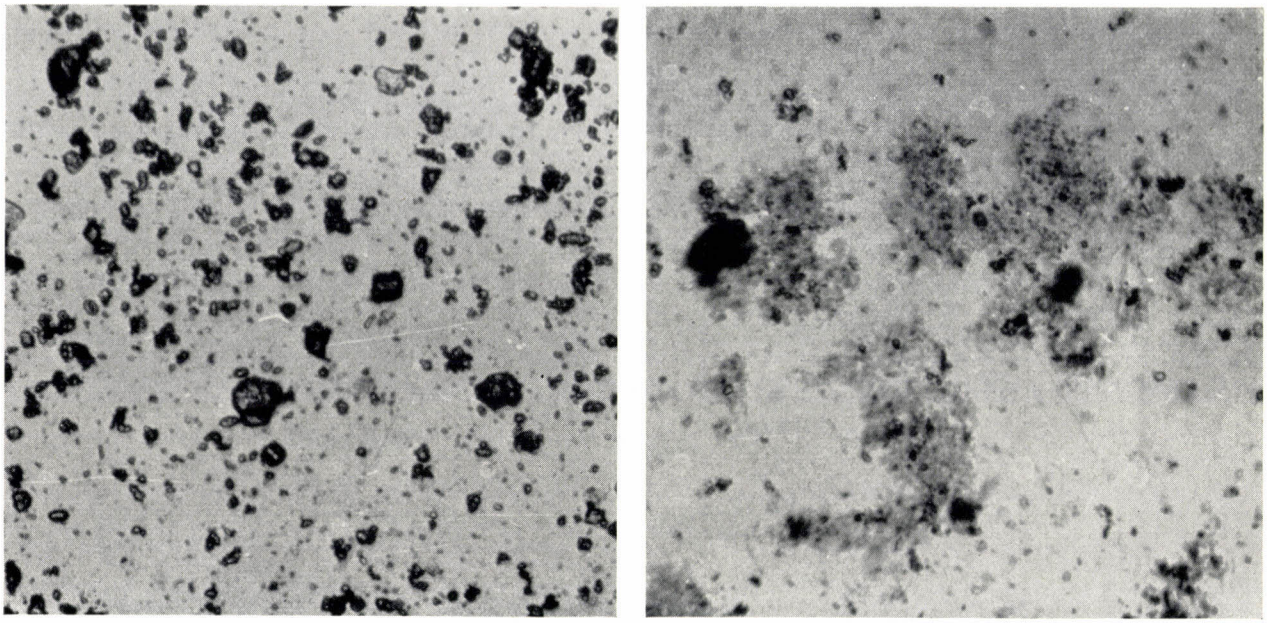
Suspensions are two-phase systems consisting of finely divided solid particles in a liquid phase (in this instance usually water), see *fig. 2a*. Like emulsions, suspensions can be prepared either by diluting a concentrated paste with water, or by mixing a wettable powder with water.

In formulating pastes the difficulty arises that the particles of the active substance must be very small, since otherwise the substance would quickly settle out when suspended in water (according to Stokes' law the rate at which such particles settle out is proportional to the square of their mean diameter). The grinding of the active substance in colloid mills is not very effective and very costly. Most insecticides and acaricides are not therefore marketed as pastes. The active ingredients for various fungicides can, however, be prepared as a precipitate in an aqueous phase; the particles so obtained are small enough (a few μ or even less) to give suitable pastes without grinding. Very good preparations can be made in this way, e.g. from basic copper salts. A disadvantage of concentrated pastes is that they require rather costly packing. Tins are out of the question because of corrosion; even coated tins are insufficiently reliable, so that all-plastic packing is indicated.

The wettable powders are more practical and cheaper than the pastes. With modern mill equipment, solid active substances together with the accessory substances can be reduced to a mean particle size of 10 μ and even smaller. Modern mills for this purpose, *fig. 3*, have a capacity of 500 to 1000 kg per hour. The contents of active ingredient can be appreciably higher than in pastes, and there are no problems in packing the powder. Nevertheless the formulation of wettable powders calls for special attention. For, in suspension the particles show a tendency to flocculate (fig. 2*b*), especially in the presence of Ca and Mg ions in hard water; the flocculated particles behave like coarser particles, i.e. they sink more quickly. The use of an excess of wetting agents, sequestering agents and dispersing agents leads to some improvement in this respect, but results in poorer adhesion to the plant; and that means little residue, especially in rainy weather.

If a suspension with a good suspensibility is to be obtained, it is essential that the active substance should be well adsorbed by the carrier. Active substances (except for lindane, which can be prepared in a chemically pure state) always contain impurities, these being frequently of an oily nature. Such impurities strongly promote flocculation and complicate the grinding process by greasing up the mills. Such substances as aluminium silicate, synthetic calcium silicate and colloidal silicic acid have been found to make very good carriers. They excel by reason of their fine structure, light weight and their high capacity for adsorbing sticky pesticides. They are fairly expensive, but this disadvantage is to some extent compensated by their strongly hydrophilic nature, which means that a low wetting and dispersing agent content will suffice.

Because the wettable powders have a high content of active substance, freight costs are relatively low. Another advantage is that, by dilution of the wett-



a *b*

Fig. 2. "Tedion V18" in suspension, *a*) not flocculated, *b*) partly flocculated. Magnification 500×.

able powder with carriers available in the country of destination, dusting powders, which are dusted dry on the plant (see the following section), can be prepared. At tropical temperatures there is a danger of lumping, especially in the case of BHC and DDT preparations containing 75% or more of the active ingredient; this can be remedied by compounding colloidal silicic acid in the preparation.

Another advantage of wettable powders should be added to those mentioned above, namely that they are in general less toxic to man, domestic animals and plants than the emulsions, and can be more readily combined with other preparations. It is therefore to be expected that emulsifiable concentrates will be used less and less and that the wettable powders will gain ground.

Dust preparations

Dusting agents are dry powders which are ready for use right away. The content of active ingredient varies from a few tenths of a per cent to a few per cent. Owing to the large bulk of inert carrier, it is of great importance that the active ingredient should be fine and evenly distributed throughout the whole mass.

The active ingredient may be solid or fluid. In the former case the dusting powder is usually prepared by mixing a finely ground pre-concentrate of the active ingredient in the carrier. The choice of the carrier and any other auxiliary substances is mainly determined by the physical properties of the active substance, such as its hardness, melting point, hydrophilic or hydrophobic nature. To ensure that the product does not lose its uniformity during storage and transport or during application, the particles must be of a uniform size; this reduces the effect of large differences in the specific gravities of the various components.

Where the active substance is more or less fluid (nicotine, parathion, malathion, toxaphene, etc), it must usually be applied on to the carrier in the form of an atomized mist (if necessary, having first been diluted with a volatile solvent). This procedure calls for good adsorbing properties on the part of the carrier, especially when pre-concentrates are to be prepared.

The finished product, irrespective of whether it has been prepared by the milling or the atomizing technique, must retain its original consistency. It must not be subject therefore, to aggregation during transport and storage when it may be exposed to pressure and elevated temperature. Aggregation does not usually give any trouble in preparations having a low content of active substance. However, the problems with pre-concentrates containing 80% or more of active substance are numerous. If the active substance has a fairly high vapour pressure (BHC, lindane), the preparation begins to agglomerate markedly at heightened temperature, to give lumps. Auxiliary substances

(conditioners) may be necessary to combat this.

Depending upon the plants that are to be treated, heavy or light dusting powders are used: heavy for low crops, light for trees.

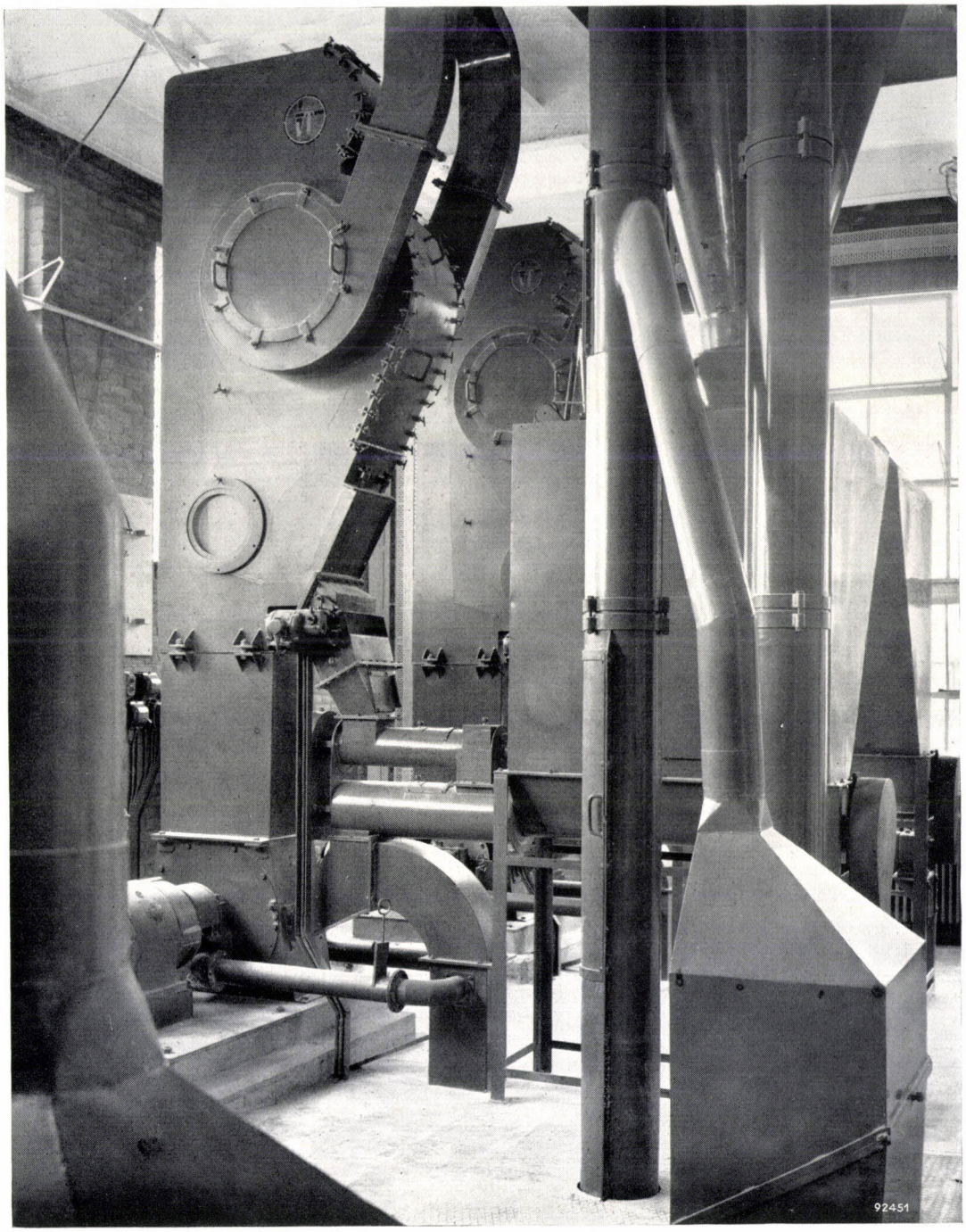


Fig. 3. One of the milling installations in the Amsterdam pesticide factory of N.V. Philips-Roxane.

The most important carriers are: magnesium and aluminium silicate (talc, kaolin), carbonates (dolomite, chalk), materials containing silicic acid (diatomaceous earth), gypsum, sulphur; further, in countries where they can be cheaply obtained: organic materials, such as wheat flour, walnut flour, sawdust and tobacco dust. Some carriers are themselves lethal to insects, either extracting water from the insect, or like crude silicic acid, damaging their cuticula by their abrasive action. It has already been mentioned in the introduction, however, that an abrasive action is bad both for the plant and for the dusting apparatus.

Many pesticides are sensitive to alkalis, thus the carrier should not have an alkaline reaction. Even if it is neutral, a compatibility test should still be

made, since certain impurities bring about catalytic decomposition of the active substance.

Aerosols

Aerosols are systems in which the solid or liquid active substance is extremely finely dispersed throughout a gaseous phase (in the present case, air) and remains suspended in the gas ("micro-mist"). In time the particles become greater by taking up moisture or by aggregation. Sedimentation sets in, so that the active substance settles on the crop. The active substance may work as a respiratory or as a contact poison. In the former case it is most effective in its micro-mist form. The contact action is only satisfactory if the particles are not too small and if the insects are very mobile, e.g. flies, which can take up a fair amount of poison during their flight and die mainly as a result of the contact action.

Aerosols, as has already been remarked, are on the market as fumigation agents, as liquid atomization preparations and as vapour agents.

Smoke generators

Smoke generators are preparations in which the active substance is mixed with an oxidizing agent and with a carrier (usually organic). Once ignited, the mixture gives off a great deal of smoke (*fig. 4*) consisting of the active substance and reaction products such as water vapour and carbon dioxide. Such preparations are effective only in confined spaces (greenhouses).

The preparation of smoke generators is not without danger, since it involves powerful oxidizing agents such as ammonium nitrate, chlorates, perchlorates or bichromates. It is absolutely essential here that all the components should be mutually compatible, if sudden undesirable reactions are to be avoided. Apart from the choice of oxidizing agents it is also important that the combustion temperature should not be too high; otherwise much of the active substance would be lost by burning. The loss of active substance is nevertheless as much as 20 to 30% even in efficient DDT and lindane smoke-generating preparations.

Liquid preparations for atomization

If a liquid is sprayed at a sufficiently high temperature and pressure a mist is obtained in which the droplets are fine enough to remain suspended in the air. This is another method used for spreading pesticides. The liquid used is a solution of an active substance in an organic solvent. Depending upon the size of the droplets, a blue (very fine) or a white (coarser) "smoke" is produced. The atomization is often effected nowadays using simply constructed apparatus working with steam. Injurious insects on dense upright crops or in woods can be effectively combated by atomization from aeroplanes. An objection to this practice, however, is that birds and useful insects, like bees, frequently suffer much harm. The success of the method, moreover, is highly dependent upon the wind.

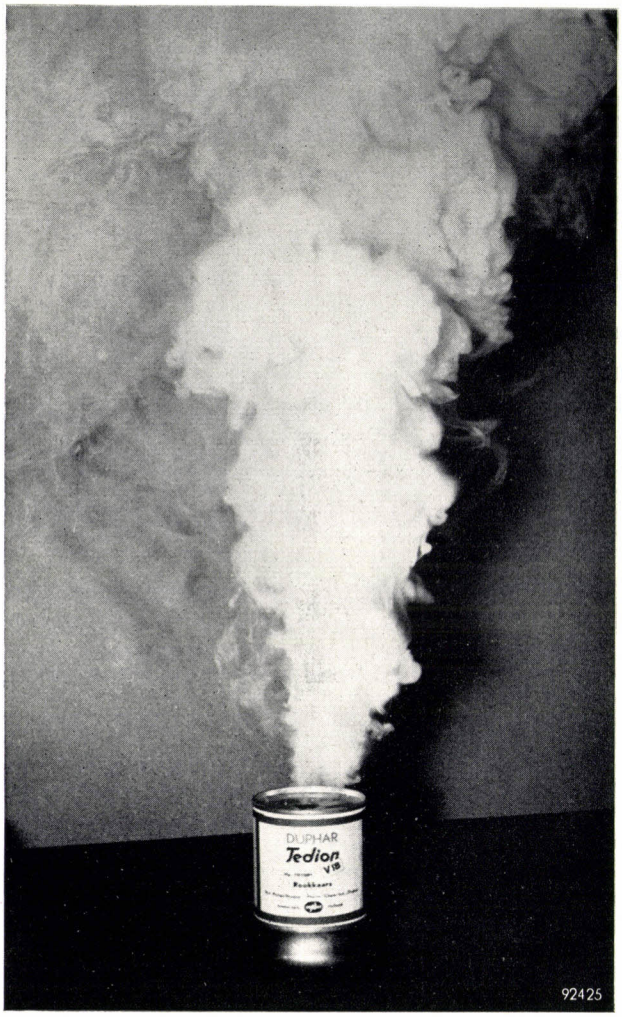


Fig. 4. Smoke evolved by smoke generator.

The solvent must have a high specific gravity (so that the mist does not settle too slowly); further, it should not burn readily and should have a low phytotoxicity, since the amount of solvent in the mist is quite considerable.

Fumigants

Fumigants are preparations containing an active substance with a high vapour pressure; the active substance is distributed by vaporization which

occurs at relatively low temperatures. The vapour molecules are imagined to combine to form a mist; this is why fumigants are classed as aerosols. Formulation is necessary only where the substance is too volatile or insufficiently volatile for normal use; the required rate of evaporation can be obtained by means of suitable additives.

A well-known vapour agent for domestic use is paradichlorobenzene, for combating moths in wardrobes. Like orthodichlorobenzene, this substance can be of good service as a soil disinfectant for exterminating termites. More widely used is "DD mixture", a mixture of dichloropropane and dichloropropylene. This is a liquid with a boiling point around 100 °C; the highly penetrating vapour is especially active *in* the soil.

Concluding remarks

The requirements with which the quality of agro-biological preparations must comply are becoming ever more exacting, and this calls for ceaseless work on their formulation. Amoung the fundamental problems that are engaging the attention of the Philips-Roxane formulation laboratory, are the flocculation phenomena and the stabilization of preparations at high temperature; these problems must be solved for each substance individually.

Naturally these physico-chemical questions must be studied with reference to the biological properties of the preparation. With a view to obtaining the optimum biological activity a great deal of attention is being given to formulations, especially for the pesticides discovered and developed by Philips-Roxane.

There are indications that the biological activity can be very much enhanced by careful choice the auxiliary substances such as carriers and surface-active agents. It is therefore not enough to test a preparation that is under development merely by physico-chemical methods; the development of a preparation can be regarded as complete only when that preparation has been put to exhaustive biological evaluation in the laboratory and in the field, and when the results of such evaluations are in accord with the preparation's physico-chemical properties. As more selective agents, particularly emphytic agents are discovered, the formulation of agro-biological preparations will become still more complex; for good results there will have to be close cooperation between the chemist and the biologist.

Summary. The control of injurious insects, red spiders, fungi and weeds is mainly done by means of chemicals: insecticides, acaricides, fungicides and herbicides, which come under the general heading of pesticides. An agro-biological preparation consists of a pesticide and various auxiliary substances (carrier, wetting agent, sticker, emulsifier, stabilizer, etc.). The compounding or formulation of these preparations is extremely important. According to the way in which the preparation is to be used, we differentiate between sprays, dusts and aerosols. The sprays are subdivided into solutions, emulsions (stock emulsions and emulsifiable concentrates), solubilized solutions and suspensions; the aerosols into smokes, liquid preparations for atomization, and fumigants. The optimum size of the particles of the active substance, the chemical compatibilities of the components of a preparation, phytotoxicity, lumping of powder preparations, flocculation of suspensions, the breaking-down of emulsions, the influence of changes in temperature and of the packing, are some of the questions touched upon in this article and occupying the attention of the "formulation chemist".

VOLUME 19, 1957/58, No. 6 pp. 177-208 Published 23rd December 1957

Philips Technical Review

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

DISPENSER CATHODES

I. INTRODUCTION
II. THE PRESSED CATHODE
III. THE IMPREGNATED CATHODE

Since its introduction in 1949, the L cathode has been adopted for many types of electronic tubes, including television camera tubes, disc triodes, klystrons and magnetrons. In the later development two new variants of the L cathode have been evolved which differ considerably in their construction from the original form and which will perhaps contribute towards a marked extension of the field of application for this type of cathode.

I. INTRODUCTION

by A. VENEMA.

621.385.032.213.13

The L cathode, a thermionic electron emitter, which was developed in the Philips Research Laboratories at Eindhoven, was described in this journal some years ago [1]). It possesses a number of highly attractive properties. Many details on the L cathode can be found in the article cited, but a few of its properties will be discussed here which have a special bearing on the development of two new types of cathodes to be dealt with in the subsequent articles by Hughes and Coppola and by Levi.

Investigations conducted by Rittner, Rutledge and Ahlert [2]) in the Philips Laboratories in Irvington (U.S.A.) have led to a clearer insight into the mechanism of the function of the L cathode. These investigations will receive attention in a later issue of this journal.

The L cathode has an emitting surface consisting of porous tungsten. The porous tungsten body forms one wall of a chamber whose other walls consist of non-porous (fully sintered and worked) molybdenum. During manufacture a tablet comprising a solid solution of barium and strontium carbonates is introduced into the chamber. The cathode is mounted in the valve and is then heated on the pump system whereupon the carbonates decompose and mixed crystals of barium-strontium-oxide are formed. The carbon dioxide liberated in the decomposition leaves the cathode through the porous tungsten body and is pumped off. The cathode at this stage does not yet possess good emitting characteristics. These are obtained by means of a special activation process, which begins as soon as carbon-dioxide evolution ceases, and which is carried out at a temperature of about 1250 °C. The electron emission from the porous tungsten surface, which is initially slight, shows a roughly uniform increase during the process. Usually the process is interrupted at a given moment so that the valve can be taken off the pump system. At this stage the emission has not yet reached its end value, but the rate of increase has become relatively slow. Activation is continued on a test rack until the emission has become constant. The cathode is then ready for use.

[1]) H. J. Lemmens, M. J. Jansen and R. Loosjes, Philips tech. Rev. **11**, 341, 1949/50.
[2]) E. S. Rittner, R. H. Ahlert and W. C. Rutledge, J. appl. Phys. **28**, 156, 1957 (No. 2).
W. C. Rutledge and E. S. Rittner, J. appl. Phys. **28**, 167, 1957 (No. 2).

During the decomposition of the carbonates the tungsten is at a temperature of about 1100 °C. The oxydation by the liberated carbon dioxide to which the tungsten is subject at this temperature has been found to have an adverse effect on the duration of the activation process. Impurities which may be present on other parts and on the wall of the tube, likewise can considerably increase this duration.

The emission properties of a cathode are characterized by the two parameters A and φ_0 in Richardson's formula, which reads:

$$j_s = AT^2 e^{-\frac{e\varphi_0}{kT}}; \quad A = A_0 e^{-\frac{e}{k}\frac{\Delta\varphi}{\Delta T}},$$

where

j_s = saturation current density in A/cm²,
A_0 = emission constant in (A/cm²) degree^{-2},
T = absolute temperature in °K,
e = charge on the electron in coulombs,
φ_0 = work function in volts extrapolated to 0 °K,
k = Boltzmann constant in watt sec/degree.

The emission constant A_0 is given by the expression

$$A_0 = G\,\overline{D}\,\frac{2\pi\, m\, e\, k^2}{h^3},$$

in which

G = occupation number (in most cases 2),
$\overline{D}$ = mean electron transmission coefficient of the potential barrier at the emitter surface,
m = mass of the electron in grams,
h = Planck's constant in watt sec².

For the L cathode, $A \approx 2.5$ (A/cm²) degree^{-2} and $\varphi_0 = 1.67$ V [2]). (This means that such a cathode will have a zero-field saturation current density of 4.7 A/cm² at a temperature of 1400 °K.) For tungsten $\varphi_0 = 4.54$ V.

The reduction in the work function is explained by the adsorption of barium and oxygen atoms on the tungsten surface. The barium atoms are produced through reduction of the barium oxide, vaporized in the chamber, by the tungsten and make their way to the cathode surface through the porous tungsten body. The special properties of the L cathode are connected with the *metallic* nature of the emitting surface and with the *dispensation* of barium from the chamber (hence the name: dispenser cathode); the dispensation process continues until all the available barium oxide has been reduced.

Among the variable characteristics of the L cathode may be listed:

1) High emission, even in continuous operation.
2) Mechanical robustness, as the tungsten surface is particularly strong.
3) The emitting surface can be rendered very flat and smooth.
4) A negligible potential drop across the cathode.
5) Long life amounting to several thousand hours, even when there is a relatively high pressure of residual gas in the tube.
6) Recoverability from poisoning.
7) Capability of withstanding strong ionic bombardment.

The first four properties derive from the metallic nature of the cathode. The current delivered by the cathode produces only a small amount of heat in the tungsten body itself, since the latter has little electrical resistance. In this respect it differs greatly from semiconductor cathodes, such as the oxide cathode. With the latter the heat generated in the cathode itself is so great at high emission current density as to cause serious damage.

The properties 5) and 6) result from the continuous dispensation of barium. Should the cathode be poisoned, it is found to recover in a period lasting from a few seconds to a few minutes, the time depending upon the temperature.

The seventh property is connected both with the metallic nature of the cathode and with the barium dispensation.

Some limitations of the L cathode, however, should not pass unmentioned:

1) The time spent on the pumping system and the activation time are fairly long. Mention has already been made of the oxidation of the porous tungsten and the effect of this oxidation on the activation time. In order to keep the oxidation within reasonable limits, the decomposition of the carbonates must occur at relatively low temperatures, which makes it a relatively slow process. Since very slow decomposition would demand far too much time, a compromise must be found.
2) A good seal between the porous tungsten body and the molybdenum is essential. If the joint were to allow barium to escape through it, instead of through the pores in the tungsten, excessive evaporation and a shorter life would result and under unfavourable conditions the emission might be affected. A perfectly tight joint can be made with resistance or arc welding. This welding operation, however, which must be carried out in a controlled atmosphere, is not very well suited to mass production.
3) The dimensions in which the L cathode can be made are subject to certain limits. It will be clear at once that the chamber construction may make for difficulties where the cathode has to be very small. On the other hand with very large cathodes, trouble may arise in the temperature

distribution over the surface, since a large part of the heat is supplied via the weld connecting the tungsten and molybdenum. The temperature is lower at a somewhat greater distance from the weld than in its immediate vicinity.

These limitations motivated an investigation which was carried out in the Philips Laboratories in Irvington. The investigation resulted in the development of two new types of cathodes in which the advantages of the L cathode were retained while its limitations were overcome. These new types, the *pressed cathode*[3]) and the *impregnated cathode*[4]) [5]) [6]), work on the same principles as the L cathode. The emitting surface is of tungsten, or a tungsten-molybdenum alloy, the work function of which has been lowered by adsorption of barium and oxygen atoms. The new cathodes differ from the L cathode in that the barium generating compound is not contained in a separate chamber but is taken up in the cavities of the porous metal body. The barium generating compound used in both cathodes produces much smaller quantities of undesirable gas than that in the L cathode. Moreover, a gas-tight weld is no longer necessary and more leeway is obtained in choosing the shape and dimensions of the cathodes.

In anticipation of the descriptions of the pressed and impregnated cathodes that are given below, it may be remarked that the emission properties of the three types are broadly the same. Two notable differences concern the method of manufacture of the various cathodes and the rate of barium dispensation. In the pressed cathode, for example, this rate of dispensation is greater than in the other types. Consequently this cathode shows a greater rate of evaporation.

The pressed cathode was principally developed as a flat cathode, the objective being to arrive at a product that would lend itself to mass production. The impregnated types can be made in any desired shape, e.g. as flat or cylindrical cathodes. The possibilities in shaping and dimensioning the latter type of cathode are very large indeed. The construction of this cathode was not primarily intended for, but does not preclude it from mass production. With mass production in mind, it is difficult at this stage to express a preference for either the one or the other cathode, since there are other properties that cannot be ignored in the assessment. Apart from the problem of barium dispensation, the problem of the insulation of the emitting body from the filament which serves to heat the cathode should be mentioned here. In its original version the impregnated cathode suffered from cathode-to-heater emission, but this difficulty has been overcome for most types of planar cathodes in a new design which is described in Part III.

When a given tube has to be fitted with a dispenser type cathode, the properties mentioned, and perhaps others besides, will determine which of the three types should be used: L cathode, pressed cathode or impregnated cathode. The choice can be expected to differ for different applications.

[3]) P. P. Coppola and R. C. Hughes, Proc. Inst. Radio Engrs. 44, 351, 1956 (No. 3).
[4]) R. Levi, J. appl. Phys. **24**, 233, 1953.
[5]) R. Levi, Le Vide **9**, No. 54, 284, 1954.
[6]) R. Levi, J. appl. Phys. **26**, 639, 1955.

II. THE PRESSED CATHODE

by R. C. HUGHES *) and P. P. COPPOLA **).

621.385.032.213.13

The electron-emitting part of the pressed cathode consists of a body which is pressed from a powdered mixture of a tungsten-molybdenum alloy and barium calcium aluminate. The latter compound is the cathode's barium source from which barium is generated as a result of reduction by the tungsten. After pressing there follows a short heat treatment above the melting point of the aluminate which is especially important to remove injurious gases still present in the pressed mass. In what follows the chemical background underlying the choice of the components will be discussed, the manufacture of the pressed cathode will be outlined, and in conclusion, its emission characteristics will be described.

Chemical considerations on the choice of the barium source and metal

From what was mentioned in the introduction it is clear that barium-strontium carbonate is not ideal as the starting material for the barium source in the

*) Philips Laboratories, Irvington-on-Hudson, N.Y., U.S.A.
**) Formerly with Philips Laboratories, Irvington-on-Hudson, N.Y., U.S.A.

L cathode. A pressed mixture of barium carbonate and tungsten does not yield any barium at all on heating and thus is not suitable for a cathode. The reaction in this case (which occurs at an appreciable rate even at 600 °C) is described by the equation [1]):

$$3\,BaCO_3 + W \rightarrow Ba_3WO_6 + 3\,CO.$$

A pressed mixture of barium oxide and tungsten powder, on the other hand, generates barium in sufficient quantity at temperatures above 1000 °C. The reaction, which is responsible also for the barium dispensation in the L cathode, proceeds according to the equation:

$$6\,BaO + W \rightarrow Ba_3WO_6 + 3\,Ba.$$

Using these substances, therefore, a cathode can be made which works on the same principle as the L cathode [2]). The electrons are derived from the tungsten whose work function has been lowered by the adsorption of a monatomic layer of oxygen and barium. There are, however, two great drawbacks associated with a cathode pressed from barium oxide and tungsten [2]):

1) The barium oxide in the cathode very readily reacts with water vapor and carbon dioxide in the air, so that assembly of the tube would have to be performed under such conditions that these gases cannot come into contact with the cathode. In practice this constitutes a serious difficulty.
2) The rate of the reaction between oxide and tungsten is inadmissibly high at the temperature desired for electron emission, which lies between 1000 °C and about 1250 °C. In consequence, the amount of barium formed per unit of time is far too great and cathodes of this construction display very rapid evaporation and have a short life.

A lower reaction rate could be obtained by substituting molybdenum for part of the tungsten (see below), but this would not solve the problem of rapid deterioration in air. Apart from the difficulties in the assembly the same difficulties with regard to reactivity with water vapor and carbon dioxide would be encountered, of course, in the manufacture of the cathodes.

A compound which is to dispense barium in combination with a suitable metal must meet the following requirements:

1) It must give off as little gas as possible. The desired temperature can then be speedily reached on the pumping system.
2) No oxygen must be present that could take part in the reaction other than the oxygen in the barium oxide.
3) High stability in air. The compound must not take up water vapor and carbon dioxide.
4) Its reactivity and rate of evaporation must be low compared with barium oxide.

The products obtained in reactions between barium oxide and other oxides, such as aluminum oxide, silicon dioxide and boron oxide, were found to come up more or less to the above requirements [2]). The most suitable was found to be the composition: 5 moles of barium oxide and 2 moles of aluminum oxide. In view of the obvious additional requirement that the pressed cathode should have as high an electron emission as possible, it has been found desirable to add an amount of calcium oxide. It was R. Levi who first showed calcium oxide to be beneficial in the impregnated cathode (see I, [6])). The way in which calcium oxide exerts this beneficial influences is not yet clear; nor is it completely known as yet how the calcium oxide modifies the properties of the barium aluminate in regard to the four above-listed requirements. Discussion of the properties in relation to these requirements will therefore be confined to the composition with 5 moles of barium oxide and 2 moles of aluminum oxide.

This formulation is an approximately eutectic mixture of the compounds $Ba_3Al_2O_6$ and $BaAl_2O_4$ [3]). Only the former adequately reacts with tungsten with generation of barium. The second compound is useful in that it increases the stability of the mixture in air and lowers the melting point. In fact, melting of the aluminate during preparation is found to be essential for satisfactory operation of the cathode. This results in the removal of most of the residual gas: accordingly, the amount of undesirable gas released later on the pump system is very small.

During operation of the cathode the tribarium aluminate may be viewed as supplying BaO by dissociation:

$$Ba_3Al_2O_6 \rightleftarrows 2\,BaO + BaAl_2O_4.$$

The barium oxide reacts further with the tungsten. The monobarium aluminate formed has no further function. Thus, no oxygen other than that in the barium oxide reacts with the tungsten, as stipulated by the second requirement.

---

[1]) R. C. Hughes, P. P. Coppola and H. T. Evans, J. appl. Phys. **23**, 635, 1952.

[2]) R. C. Hughes and P. P. Coppola, J. appl. Phys. **23**, 1261, 1952.

[3]) N. A. Toropov and F. Ya. Galakhov, Doklady Akad. Nauk S.S.S.R. **82**, 69, 1952.

The degree of stability in air of the aluminate with the composition 5 moles barium oxide and 2 moles aluminum oxide can be seen from *fig. 1*. This figure shows the time dependence of the weight increase for powdered barium oxide (*I*), powdered aluminate (*II*), a fused pellet of aluminate (*III*) and a pressed cathode (*IV*) on exposure at a temperature of 20 °C to air having a relative humidity

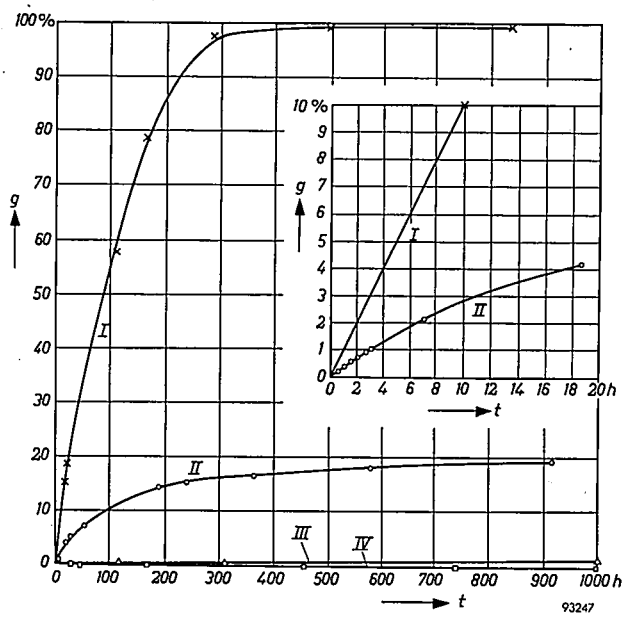


Fig. 1. The stability in air of the pressed cathode and of various emitting materials that may be used in this cathode. The percentage increase in weight g by uptake of water is plotted as a function of the time t. *I* powdered BaO, *II* powdered Ba-aluminate, *III* fused pellet of Ba-aluminate, *IV* pressed cathode.

of 45%. The plot demonstrates that powdered barium aluminate is far more resistant to water and carbon dioxide in the air than is powdered barium oxide; still better are the fused aluminate pellets, which show hardly any increase in weight. Finally, with the cathode itself, where the aluminate is interspersed between the separate tungsten grains, no increase in weight can be detected even after 1000 hours.

In conclusion, there is the question of the reactivity of the barium aluminate and the evaporation of barium oxide from this aluminate. A pressed mixture of tungsten powder and aluminate powder ($5BaO : 2Al_2O_3$) is found to react on heating in a manner not quite desirable for a cathode. True, the amount of barium that escapes from the mixture per second at a temperature of 1130 °C is less than 1/100 of that given off by an equal amount (in grams) of the mixture of tungsten and barium oxide, but even at that, it is still excessive. The rate of escape of barium is determined not only by the chemical properties of the system, but also by the porosity of the tungsten matrix. This porosity is fairly large and cannot readily be further reduced by mere compression, while high-temperature sintering detracts from the properties of the cathode. The problem must therefore be resolved chemically.

Thermodynamic considerations showed that a tungsten-molybdenum alloy could be expected to give a low rate of barium production. This was confirmed by experiments. A discussion of these considerations and experiments will now be given.

The pressure of barium vapor in equilibrium with a barium compound and a reducing metal may be calculated by thermochemical methods [4])[5]). The maximum rate at which barium will evaporate from this system can be determined from the equilibrium pressure. The reaction between the barium aluminate and the reducing metal can be considered to proceed in two stages:

1) the dissociation of the aluminate to form barium oxide, and
2) the reduction of the barium oxide to form barium.

Barium generation will be discussed first. The reaction involved here can be represented by the equation:

$$M_s + BaO_s \rightleftarrows MO_s + Ba_g,$$

where M is the reducing metal (assumed for simplicity to be bivalent); the suffixes s and g indicate whether the substance participates in the reaction in the solid or gaseous state. If only one of the reactants is gaseous, as in the above example, the equilibrium pressure can be derived from the equation:

$$\Delta G_T = -RT \ln P_{Ba} = -4.575\, T \log P_{Ba}\,.$$

Here ΔG_T is the change in free energy occurring in this reaction and is expressed in calories. R is the gas constant, and T the absolute temperature. The pressure P_{Ba} of barium is measured in atmospheres. The change in free energy can be written as:

$$\Delta G_T = \Delta H - T\Delta S.$$

For the change ΔH in enthalpy occurring in the reaction and the change ΔS in entropy a good approximation is obtained by substituting the values at room temperature, which are obtained from the tables [6]).

As regards the dissociation of tribarium aluminate ($Ba_3Al_2O_6$), this will produce barium oxide and monobarium aluminate as stated above, while the

---

[4]) A. H. White, J. appl. Phys. 20, 856, 1949.
[5]) E. S. Rittner, Philips Res. Rep. 8, 184, 1953.
[6]) F. D. Rossini, D. D. Wagman, W. H. Evans, S. Levine and I. Jaffe, Nat. Bureau of Standards, Circular 500, 1952.

monobarium aluminate ($BaAl_2O_4$) likewise may dissociate, forming barium oxide and aluminum oxide. The heat of dissociation generated in either reaction [7]) is added to the heat of reaction of BaO plus metal (W or Mo) in calculating the resulting pressure of barium in equilibrium with the reaction: aluminate plus metal. The change in entropy in the dissociation of aluminate can be neglected. From the heats of dissociation of the tribarium aluminate and monobarium aluminate, the barium oxide pressure can also be calculated.

The rates of evaporation of barium and barium oxide are calculated from the pressure at equilibrium using the well-known relation from the kinetic theory of gases:

$$Q = 5.83 \times 10^{-2} P \sqrt{\frac{M}{T}},$$

where Q represents the amount of material in grams that evaporates, calculated per cm² and per sec, P the equilibrium pressure in mm Hg, T the absolute temperature and M the molecular weight of the substance in question. Since it remains uncertain whether the equilibrium pressure will be attained in the porous substance, and since in addition the vapor may be impeded in flowing to the outside, Q will represent the largest possible quantity of material which can be obtained per sec and per cm² from the pressed cathode.

The results of the calculation for the two aluminates $BaAl_2O_4$ and $Ba_3Al_2O_6$ with the metals tungsten and molybdenum are given in *Table I*. The latter also shows the behaviour of a pressed mixture of the above substances as a cathode. It has further been assumed that the normal tungstate $BaWO_4$ or molybdate $BaMoO_4$ form. It is clear from the table that molybdenum is far less reactive than tungsten. It is likewise apparent that tungsten activated by tribarium aluminate emits, but as a cathode has only a moderate life. This is a consequence of fairly rapid evaporation as may be seen by comparing the evaporation rate with that of an L cathode at the same temperature.

Table I. Calculated rate of evaporation of barium and barium oxide for various pressed mixtures. Temperature: 1400 °K.

System	Pressure in 10^{-6} mm Hg		Rate of evaporation in μg/cm²h		Behaviour as an electron emitter
	Ba	BaO	Ba	BaO	
W + BaO	80	6	5200	420	emits, short life
Mo + BaO	1	6	65	420	emits, moderate life
W + $Ba_3Al_2O_6$	0.5	0.14	33	10	emits, moderate life
Mo + $Ba_3Al_2O_6$	0.009	0.14	0.6	10	does not emit
W + $BaAl_2O_4$	0.0008	0.0011	0.05	0.08	does not emit
Mo + $BaAl_2O_4$	0.00002	0.0011	0.001	0.08	does not emit

From the foregoing it is to be expected that evaporation will be diminished if the tungsten is alloyed with molybdenum. *Fig. 2* shows that this is indeed the case. Here the evaporation rate is plotted for pellets pressed from different W-Mo alloys with barium aluminate of the composition $5BaO:2Al_2O_3$. The graph likewise gives the calculated vapor pressures for the pure molybdenum and pure tungsten mixtures.

The figure shows that the increase in evaporation observed on adding tungsten to molybdenum is proportional to the molar concentration of tungsten. In view of the calculated values of the evaporation rate listed in Table I it can be assumed that the product of evaporation in the case of pure molybdenum consists almost entirely of barium oxide, while the increase due to the tungsten consists of

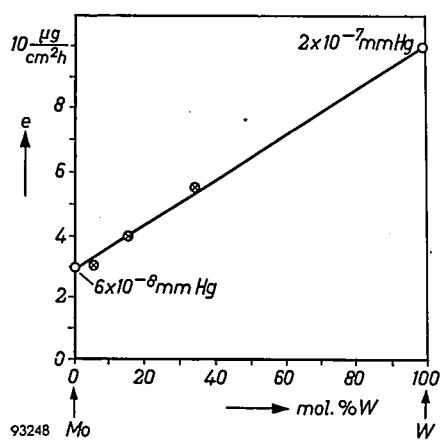


Fig. 2. The evaporation rate e of barium and barium oxide from pressed cathodes using a tungsten-molybdenum alloy, as a function of the tungsten content of the alloy.

[7]) The values of the dissociation heats are given by R. B. Peppler and E. S. Newman, J. Res. Nat. Bur. Standards 47, 439, 1951.

barium. It can easily be demonstrated theoretically that the barium generation should be proportional to the tungsten content of the tungsten-molybdenum alloy (see I [3]).

The material that was eventually chosen for the pressing of cathodes consisted of 10% by weight of barium aluminate containing barium oxide and aluminum oxide in the molecular proportion of 5 to 2, and 90% by weight of a molybdenum-tungsten alloy containing 25% by weight of tungsten. This formulation gave cathodes which showed very high stability in air, contained a minimum of gas, enabled rapid activation, and combined a high emission density with long life, while at the same time the evaporation of barium and barium oxide remained within reasonable limits. The emission density was, however, appreciably smaller than that of the L cathode. As has already been mentioned, the addition of CaO increased the emission considerably, close to the L cathode value.

The manufacture of the pressed cathode

The starting materials for the preparation of the requires aluminate are barium carbonate, calcium carbonate and aluminium oxide. These are weighed out and carefully mixed in the proportions:

$BaCO_3$	5 mol	or	$CaCO_3$	0.3	mol
Al_2O_3	2 mol		$BaCO_3$	5	mol
			Al_2O_3	2	mol

The mixture is heated in air until it fuses. After cooling, the product is reduced to a fine powder, sieved and is kept sealed off from the air.

The metal powder is mixed very carefully with the powdered aluminate in the proportions by weight of 9 : 1 and kept sealed off from the air until the time comes for pressing. This latter operation can be performed in a number of ways, depending on the required shape of the cathode and on the manner in which it is going to be fitted in the holder. Here an inexpensive method will be described which has led to good results. The method was designed for a flat cathode having a circular emitting surface with a diameter of 3 mm. In the pressing process a fixed volume of the above-described aluminate-metal mixture is introduced into a small molybdenum cylinder. The latter is mounted in a die. The mixture is then compressed with the required pressure of 11 tons/cm^2 using a suitable plunger. The compact is then taken from the die, and the cathode is heated at a temperature of 1800-1900 °C in vacuo, or better in an atmosphere of hydrogen. This is a very important stage in the manufacture of the cathode. The aluminate fuses, giving off occluded gas, whereupon capillary forces effect an even distribution of the aluminate among the grains of metal.

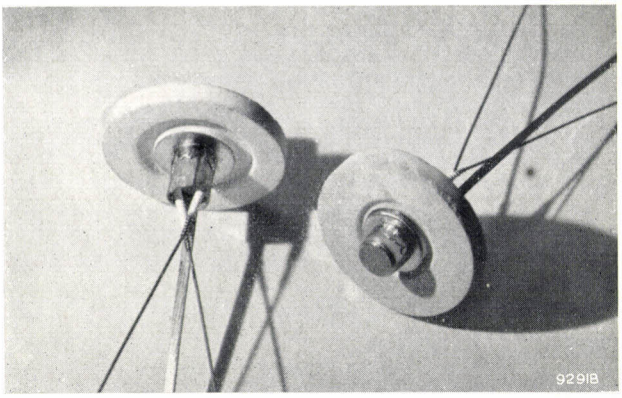


Fig. 3. Pressed cathodes with flat emitting surface (diameter 3 mm) contained in cylindrical molybdenum holders.

The molybdenum cylinder is fashioned very simply from a thin plate. This is rolled, the ends being left free. During pressing the end of the cylinder containing the powder is crimped in somewhat by the die. This will ensure a strong bonding of the pressed material to the holder after heat treatment.

The pressed surface is uniformly flat. Over the 3 mm diameter the surface was found to deviate by no more than 3 microns from a flat.

The filament for heating the cathode is fitted into the molybdenum cylinder. Cathodes of the type thus produced and intended for use in cathode-ray tubes of the grounded-cathode type are shown in *fig. 3.*

The emission characteristics

For examining the emission characteristics of these cathodes, they are mounted at a short distance from an anode in a tube. The distance between the two electrodes is made so small (a few tenths of a millimetre) that the saturation current can be drawn at a voltage of say 1000 V. The diode formed in this way is connected to a high-vacuum pump which is capable of lowering the pressure in the tube to below 10^{-6} mm Hg. The tube contains a getter to remove the gases that may be liberated after sealoff. The tube is heated for one hour in an oven at 450 °C. After cooling, the cathode is heated by applying a voltage across the heater filament, the temperature being gradually raised to 1200 °C in about 5 minutes. The anode is then degassed by high-frequency heating, and a potential of 300 V applied between cathode and anode. A current is found to flow which increases rapidly and in about 1 minute has attained a constant value. The getter

is then activated and the tube is removed from the pump system by sealing the glass stem. At this point the saturation current density is already roughly half the maximum value, which is attained after the brief aging that now follows.

The saturation current cannot be measured with a constant voltage of say 1000 V, as this would lead to excessive heating of the anode. For this reason, a pulse voltage obtained from the discharge of a condenser is applied to the anode (frequency 60 per second, pulse length ~ 100 μsec). At the same time the voltage across the diode and the current passing through the diode are reproduced on an oscilloscope in a horizontal and a vertical direction respectively [8]). *Fig. 4* shows the relationship between voltage and current for various cathode temperatures (the current density rather than the current is plotted on the vertical axis). The cathode used in this experiment contained Ba-Ca-aluminate.

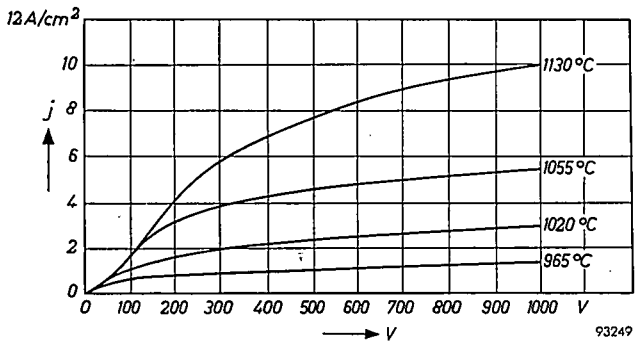


Fig. 4. Current density j of a pressed cathode (with CaO), measured in a diode and plotted as a function of the applied anode voltage V at various cathode temperatures.

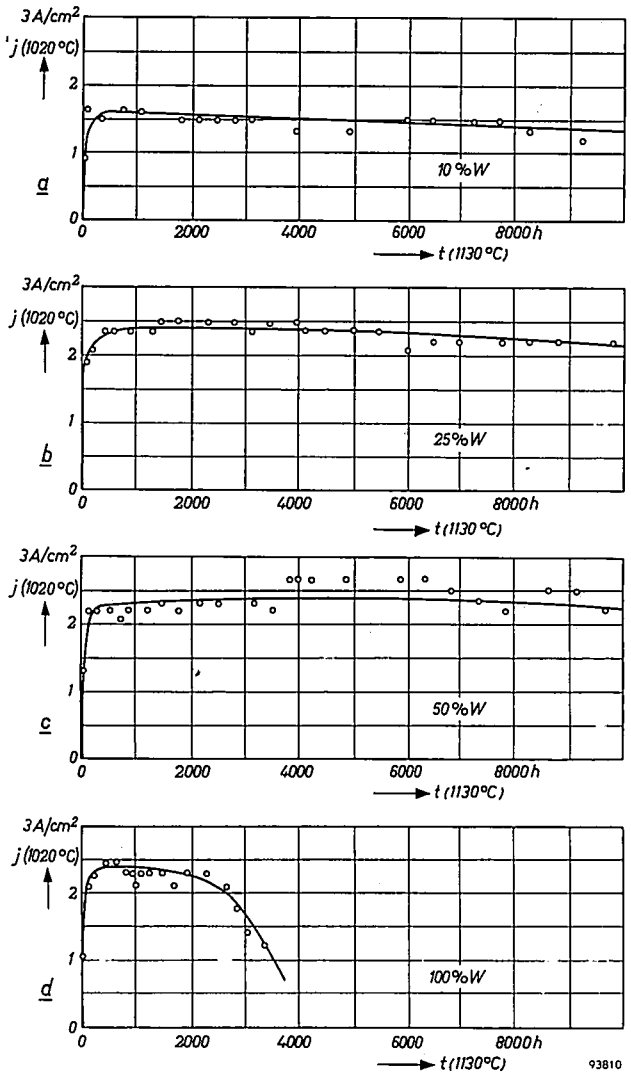


Fig. 5. Current densities j of various pressed cathodes at an anode potential of 1000 V, as a function of the operation time t.
a) Alloy of 10% W — rest Mo (by weight).
b) 25% W — rest Mo.
c) 50% W — rest Mo.
d) Pure W.
All four cathodes were made using barium aluminate ($5BaO : 2Al_2O_3$). During the hours in operation, plotted along the abscissa, the cathode temperature was 1130 °C; the temperature was reduced to 1020 °C for each measurement.

As far as it has been possible to ascertain (within the limits of tube loading studied), the currents that flow through the diode for a D.C. potential are the same as those obtained with a pulse voltage as plotted in fig. 4.

In *fig. 5* the current density measured at a cathode temperature of 1020 °C and an anode voltage of 1000 V, is plotted as a function of the time in operation. The various curves relate to cathodes pressed from barium aluminate ($5\ BaO : 2\ Al_2O_3$) and various Mo-W alloys. With pure tungsten the life is too short, owing to excessive evaporation. As the molybdenum content increases, evaporation diminishes and the life is prolonged. At first the current density remains the same, but at a high molybdenum content, the emission falls, becoming negligible for 100% Mo. The alloy containing 75% by weight of molybdenum and 25% by weight of tungsten as mentioned above is about the best compromise.

The improvement in the emission found on adding calcium oxide can be seen from *fig. 6*. The temperature was 1130 °C.

Determination of the constants in Richardson's formula give the values:

$\varphi_0 = 1.7$ V and $A = 1.8$ (A/cm²) degree^{-2} for the barium aluminate, and

$\varphi_0 = 1.7$ V and $A = 2.4$ (A/cm²) degree^{-2} for the barium-calcium aluminate.

The activation process can be visualized by producing an electron-optical image of the cathode on

[8]) R. Loosjes and H. J. Vink, Philips Res. Rep. 2, 190, 1947.

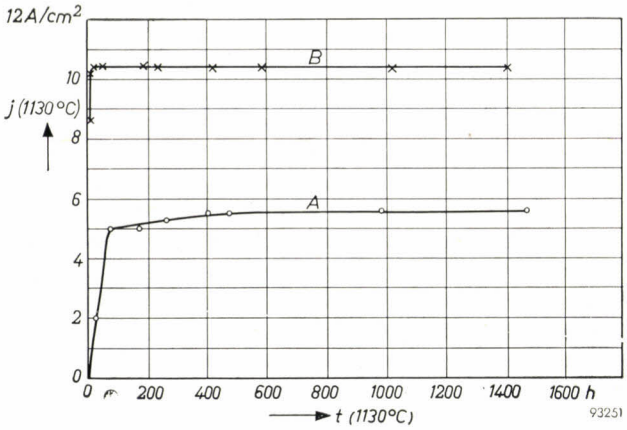


Fig. 6. Current densities of pressed cathodes at an anode voltage of 1000 V, as a function of the life. Temperature was the same during operation and measurement, namely 1130 °C. *A* barium aluminate; *B* barium-calcium aluminate.

a fluorescent screen. The result for a temperature of 950 °C is reproduced in *fig. 7*. Naturally at this temperature the activation process proceeds far more slowly than at the usual temperature of 1200 °C.

Experiments with pressed cathodes, carried out with a tube which had been exhausted on the pump to 10^{-3} mm Hg and whose pressure after sealing had been further reduced to about 10^{-6} mm Hg with the aid of a getter, showed the saturation current density to be pretty much the same as that attained under very much more favorable conditions on the pump system. It is to be expected, therefore, that pressed cathodes will lend themselves for use in tubes exhausted with two-stage rotary pumps which can get down to a pressure of 10^{-3} mm Hg and which are usual in mass production equipment.

Exposure of the activated cathode to the air will necessitate a second activation, but it has been found that this does not take more time than the first, and the current density obtained is the same as before admission of the air. This process can be repeated five times before its effect on the activation time and emission becomes appreciable.

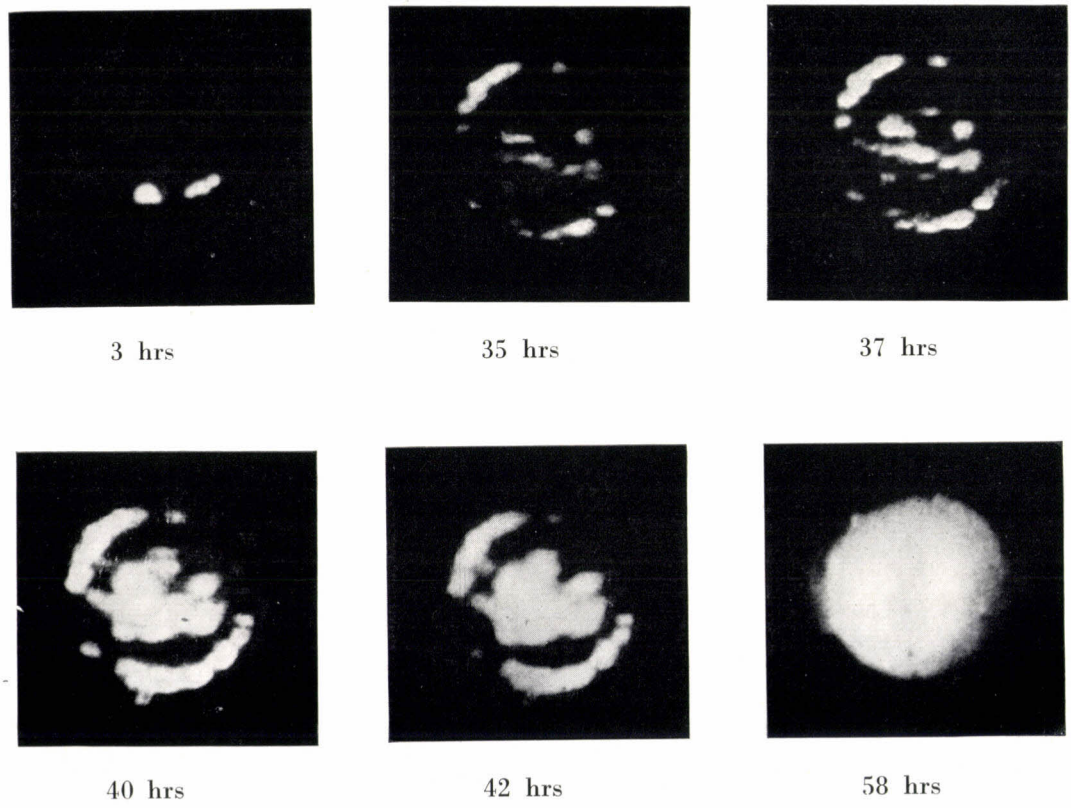


Fig. 7. Electron-optical images of the emitting cathode on a fluorescent screen, photographed at various phases in the activation process. (The latter was performed at 950 °C and was therefore much slower than normal.) The activation time is indicated under each photograph.

III. THE IMPREGNATED CATHODE

by R. LEVI *).

621.385.032.213.13

The impregnated cathode is made by impregnating a porous tungsten body with a suitable barium compound in the molten state, this body having one or more surfaces that are to emit electrons at high temperatures. The most salient feature of this cathode is the wide variety of sizes and shapes made possible by this method of fabrication.

In an early version of the cathode, the impregnant consisted of a mixture or normal and basic barium aluminates. While this cathode displayed the desired improvements in processing and manufacture over the L cathode, the emission was significantly lower. A later version of the cathode, which forms the subject of this present article, overcame this emission difficulty by the addition of calcium oxide.

Choice of the barium source and the metal

In the L cathode the activating material, barium and barium oxide, makes its way to the emitting surface through the pores in the porous tungsten body. In so doing, this stream of atoms and molecules encounters a certain resistance which depends upon the size of the pores. This resistance will generally be greater, the smaller the pores. Thus the dispensation and also the evaporation of barium and barium oxide can be controlled within certain limits by controlling the porosity. In the pressed cathode, it was not possible to make the pores small enough to prevent excessive evaporation. The problem of reducing the barium dispensation therefore had to be approached in a different way. A solution was found, as described in Part II of this article, by using a molybdenum-tungsten alloy instead of pure tungsten. Such an artifice is not necessary for the impregnated cathode. The latter is made of tungsten which is given a porosity of only 17% by the choice of a suitable pressure and sintering temperature during fabrication from tungsten powder. In other words, the tungsten has a density of 83 % of the bulk value; the remaining volume consists of pores between the grains. (In the L cathode the porosity is usually 27%, and in the pressed cathode even as much as 40%.) The use of this low-porosity tungsten in conjunction with barium aluminate ($5BaO : 2Al_2O_3$) and, better still, with the composition $5BaO.3CaO.2Al_2O_3$, has ensured a suitably limited dispensation of barium and barium oxide.

Barium aluminate was discussed in connection with the pressed cathode. What was said regarding its stability to air, slight release of gas, low reactivity and absence of excess oxygen, applies equally well when this material is used in the impregnated cathode. Addition of calcium oxide raises the emission and reduces evaporation.

Manufacture of the impregnated cathode

The aluminate is prepared in the manner described for the pressed cathode, but for the impregnated cathode, a different calcium-oxide-containing composition has been chosen. This is prepared from a mixture, comprising:

$CaCO_3$	3 moles
$BaCO_3$	5 moles
Al_2O_3	2 moles.

Actually the amount of $CaCO_3$ may vary over broad limits, the lower limit being 0.3 mole.

Usually the aluminate is prepared as pellets by melting, these being kept closed off from the air and powdered shortly before use.

The porous tungsten body, one or more surfaces of which are to act as sources of electrons, is first fashioned to the required dimensions. At the same time the surfaces, where this is necessary, are given an appropriate finish. Fashioning and finishing are easily done by means of a known technique developed by the author and described in this Review some time ago [1]). This technique consists in first filling the pores of a tungsten ingot, which has been sintered to the required density, with a suitable material, commonly copper. In this state, the tungsten can be turned, drilled, milled and polished. When the final shape is obtained, the copper is removed, e.g. by evaporation, without any dimensional change.

With 17% porosity, almost all the pores are found to be interconnecting, so that the porous body can be nearly completely filled with copper. The porosity could be made smaller, and this might be an advantage as regards the use of the porous body as a cathode, since it would reduce the evaporation of barium and barium oxide. However, on further reducing the porosity beyond this value, the machining becomes progressively more difficult because of a rapid increase in the percentage of non-connecting pores which the copper cannot infiltrate.

*) Philips Laboratories, Irvington-on-Hudson, N.Y., U.S.A.

[1]) R. Levi, A technique for machining tungsten, Philips tech. Rev. 17, 97-100, 1955/56.

The way in which the cathode is impregnated with aluminate after machining depends upon its shape and size. In general the powdered aluminate is brought directly in contact with the body, and the temperature is raised above the melting point of the aluminate. This is usually carried out in a tungsten or molybdenum furnace, heated by a directly applied current or by induced high-frequency currents. In some cases it is possible to raise the body itself to a sufficiently high temperature by one of these methods of heating. The molten aluminate is then sucked up into the porous tungsten by capillary forces. Excess aluminate can easily be removed from the tungsten surface after cooling. Initially the impregnation process was carried out in vacuo; it was later found advantageous to carry out this operation in hydrogen, since at 1 atm of hydrogen the impregnation time becomes much less critical and the activation time is radically shortened, viz. from about 70 to 3-6 hours in the case of the impregnant containing CaO.

Before the cathode can be used in a tube, the impregnated tungsten body must be attached in an appropriate manner to a holder. Molybdenum is about the best material for the holder: substances that exert a strong reducing action on the aluminate are ruled out, while, moreover, the material must itself evaporate to only a negligible extent at the operating temperatures of the cathode: roughly 1000 °C to 1300 °C. The heater filament which brings the cathode to the desired temperature also has to be accommodated in the holder. The heater filament is made of tungsten and is usually insulated from the holder by means of aluminum oxide.

The method of fitting the tungsten body into the holder depends upon the application. Let us first consider a flat cathode of circular cross-section. In many cases it will suffice if the tungsten disc, which is slightly tapered at the emitting end, is fixed in place in a molybdenum cylinder with an inner supporting ridge by spinning the molybdenum against the tapered part, *fig. 1*. The heater filament can then be pushed into the molybdenum cylinder behind the emitting disc. An advantage of this construction is the good heat transfer from filament to emitting surface.

This method of assembly cannot be used for flat cathodes employed in, say, television picture tubes. In such tubes there may be a potential difference between the heater filament and the cathode part proper, and thus, to ensure that the electrons emitted do not reach the heater filament when the latter is positive with respect to the actual cathode, the two must be completely separated from each other. This would indeed be desirable in any case, since the evaporation of barium or barium oxide towards the heater filament in these tubes may lead to undesirable effects.

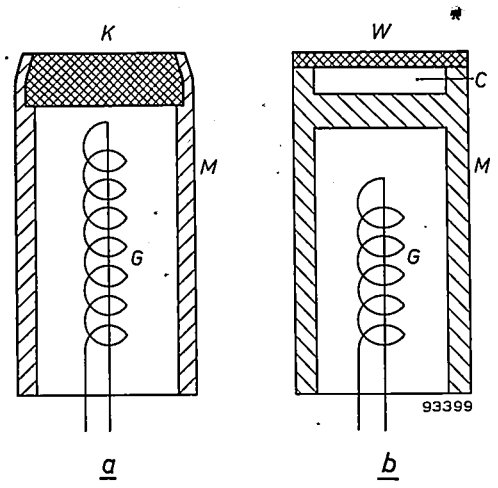


Fig. 1. *a*) A simple way of fitting a flat impregnated cathode (*K*) in a molybdenum holder. The latter consists of a cylinder (*M*) with a supporting ridge, in which the cathode disc, i.e. the porous aluminate-impregnated tungsten body, is fixed by bending over the lip of the cylinder. The heater filament (*G*) is pushed into the cylinder.
b) The L-cathode construction is shown for comparison. The porous tungsten body *W* forms one wall of the chamber *C* which contains a reservoir of a suitable barium compound.

Complete separation of filament and cathode is obtained in the following method for preparing the flat cathode.

In addition to the impregnated tungsten disc and the molybdenum cylinder (the latter can be made from a plate which is bent into shape leaving a very narrow seam), a small molybdenum plate is required. This plate is somewhat concave in shape. The three parts are fitted together as shown in *fig. 2* and are then resistance-welded together in one operation. The weld need not meet very exacting demands. It must be mechanically strong, but unlike the weld in the L cathode, it need not be a perfect seal, provided the plate is made a few tenths of a millimetre bigger in diameter than the molybdenum cylinder. The reason is that barium and barium oxide are found to be able to move only a certain distance over the surface of the metal before they again evaporate. At the temperatures employed here, this distance, as work by Rittner and coworkers has shown, amounts to a few tenths of a millimetre for tungsten [2]) and molybdenum [3]). If the smallest distance across the molybdenum is

[2]) E. S. Rittner, R. H. Ahlert and W. C. Rutledge, J. appl. Phys. 28, 156, 1957 (No. 2).
[3]) E. S. Rittner and R. H. Ahlert, submitted for publication in J. appl. Phys.

made bigger than this migration distance, then the barium, which migrates over the molybdenum from the sides of the tungsten body, cannot reach the heater filament, even if the weld between the molybdenum disc and the cylinder is not a perfect seal.

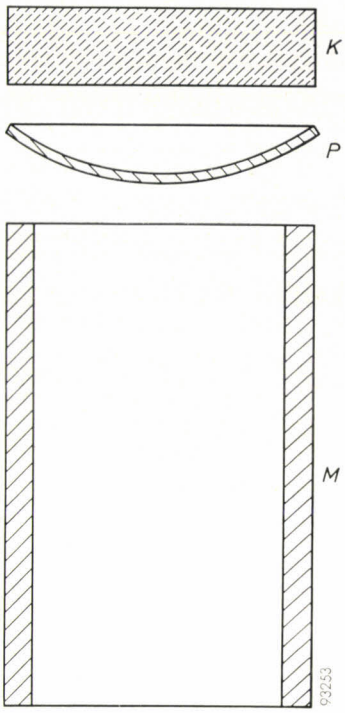


Fig. 2. The construction of a flat impregnated cathode in which the heater filament and the emitting surface are completely separated from each other. The cathode disc K, the molybdenum cylinder M and the molybdenum separating plate P are resistance-welded together in one operation.

If the emitting surface is not flat, but e.g. concave, similar considerations apply. Cathodes of this type are displayed in *fig. 3*.

In cylindrical cathodes the emitting tungsten body is again made first. This is done in the manner

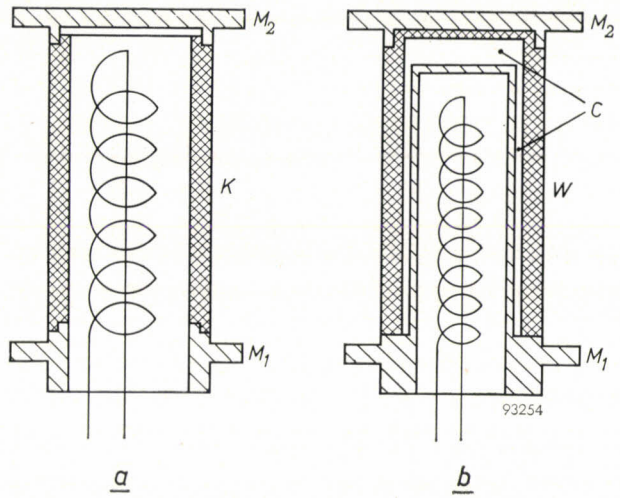


Fig. 4. *a*) Fixing a cylindrical impregnated cathode (K) into a molybdenum holder (M_1, M_2).
b) The corresponding construction of the cylindrical L cathode, shown for comparison. The chamber C is charged with the barium compound.

already described, i.e. by machining the porous tungsten after it has been impregnated with copper. This body can now be fixed into the holder in several ways, *fig. 4* showing a possible construction. On comparison with the corresponding construction for the L cathode it will be obvious that the impregnated cathode is easier to make, especially for small cathodes, and that more space is available for the heater.

It was recently discovered that tungsten, sintered to the required high density, can still be machined, e.g. by turning, drilling, etc., after it has been impregnated with the aluminate. Fairly large pieces of cathode material can be made in ingot form which are fashioned later into the desired shape and dimensions. This simplification in cathode manufacture is very attractive as it reduces the number of steps

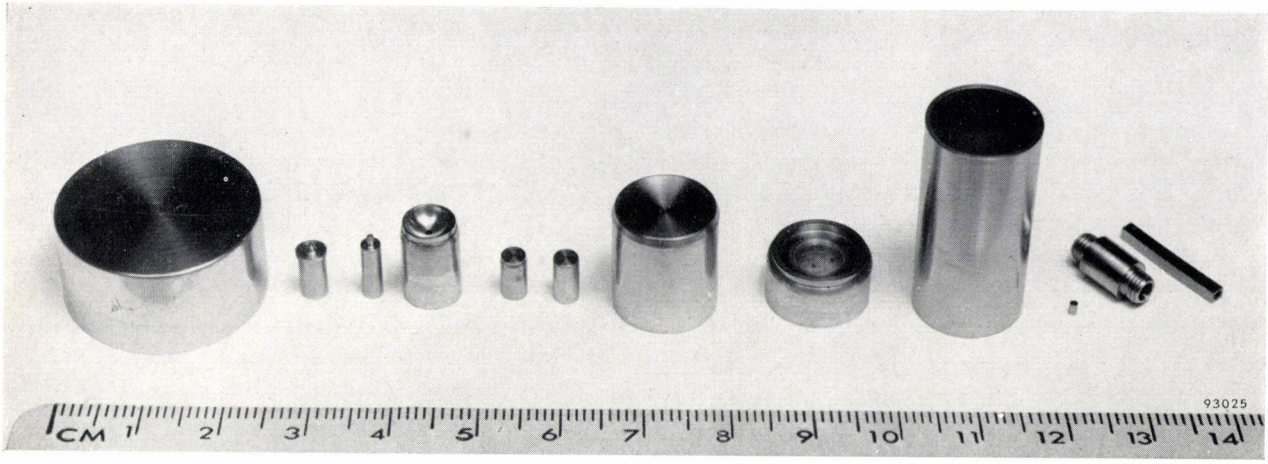


Fig. 3. A number of impregnated cathodes with flat, concave or cylindrical emitting surfaces. It is clearly seen that this type of cathode can be made in many different shapes and sizes.

required and eliminates the need for expensive vacuum equipment. The machining of the ingots impregnated with the aluminates is less easy than when the tungsten is impregnated with copper and therefore this technique cannot be used in the case of the most complicated shapes. In the case of the simplest cathode shapes, such as those of the planar type with a flat emitting area of moderate size, the emissive body can, of course, also be made by compressing the tungsten powder directly into a die, sintering to a porosity of 17% and subsequently impregnating with the aluminates.

The greater latitude allowed in design is the main feature of the impregnated cathode as compared with the pressed cathode. In addition the former exhibits a smaller rate of evaporation.

The emission characteristics

The emission characteristics are studied in a flat diode with a sufficiently small distance between cathode and anode, as described for the pressed cathode (II, page 183). The tube is heated for 1 hour in an oven at 450 °C and a current is then passed through the heater filament, the cathode temperature being gradually raised to 1250 °C. If the cathode is to operate at a higher temperature, the cathode should attain a temperature during this phase which is at least 50 °C higher than the operating temperature.

After the gases released have been pumped off and the pressure in the tube has fallen to its original value, the anode is heated by means of a high-frequency induction current. A direct voltage is then applied between cathode and anode, the cathode being at a temperature of 1250 °C. A current is found to pass which attains a maximum value within a few minutes. The getter is then activated, whereafter the tube can be removed from the pump. Complete activation is then attained within a few hours by heating the cathode at 1225 °C. It is not necessary to apply an anode voltage during this process.

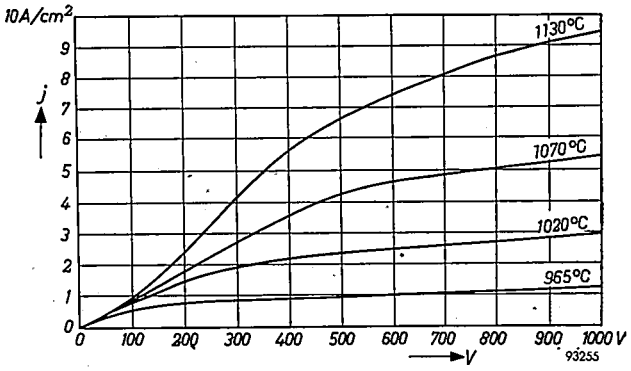


Fig. 5. Current density j of an impregnated cathode measured in a diode as a function of the anode voltage V, for various cathode temperatures.

The emission is measured also in this case by using voltage pulses of about 100 μsec duration. *Fig. 5* shows how the currents so obtained depend upon the voltage applied. The current density is plotted along the ordinate. *Fig. 6* shows that the addition of calcium oxide leads to an appreciable rise in the emission. The values of φ_0 and A in Richardson's equation are calculated from the current density found, yielding in this case $\varphi_0 = 1.67$ V and $A = 2.5$ (A/cm²) degree^{-2} [4]).

If an activated cathode comes in contact with the air, it loses its emission. It is found however that re-activation carried out as described above, restores the cathode to the same emission as before and the time required for re-activation is no longer than in the first activation. This process can be repeated several times before any effect on the re-activation time and emission current can be detected.

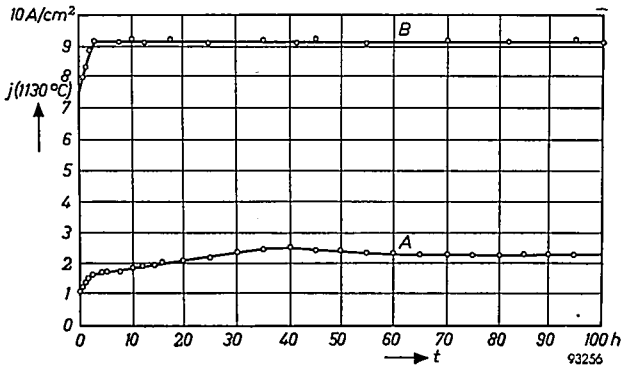


Fig. 6. The improvement in the emission of the impregnated cathode on replacing the barium aluminate (curve A) by barium-calcium aluminate (curve B). The graph shows the rise in current density during activation subsequent to seal-off.

Evaporation and life

In contrast to the situation in the L cathode, the rate of evaporation from the impregnated cathode is strongly time-dependent. Almost all of the total barium evaporates at a rate given by a $t^{-\frac{1}{2}}$ law [5]). While the high initial rate of evaporation can be harmful in certain applications, it has a twofold advantage of providing rapid activation and of overcoming poisoning which is usually most serious during early stages of life. The temperature-dependence of the evaporation follows an $\exp(-Q/kT)$ law within the normal operational range of temperatures; Q is the heat of the reaction producing barium, and may be estimated to be 4 eV per atom.

[4]) The value of A computed from the figures quoted in I [6]) is 20% higher. Further work of Rittner, Rutledge and Ahlert [5]) has shown that the effective emitting area is 20% greater than the area of the tungsten emitting surface, due to the fact that part of the molybdenum support emits too.

[5]) E. S. Rittner, W. C. Rutledge and R. H. Ahlert, to be published.

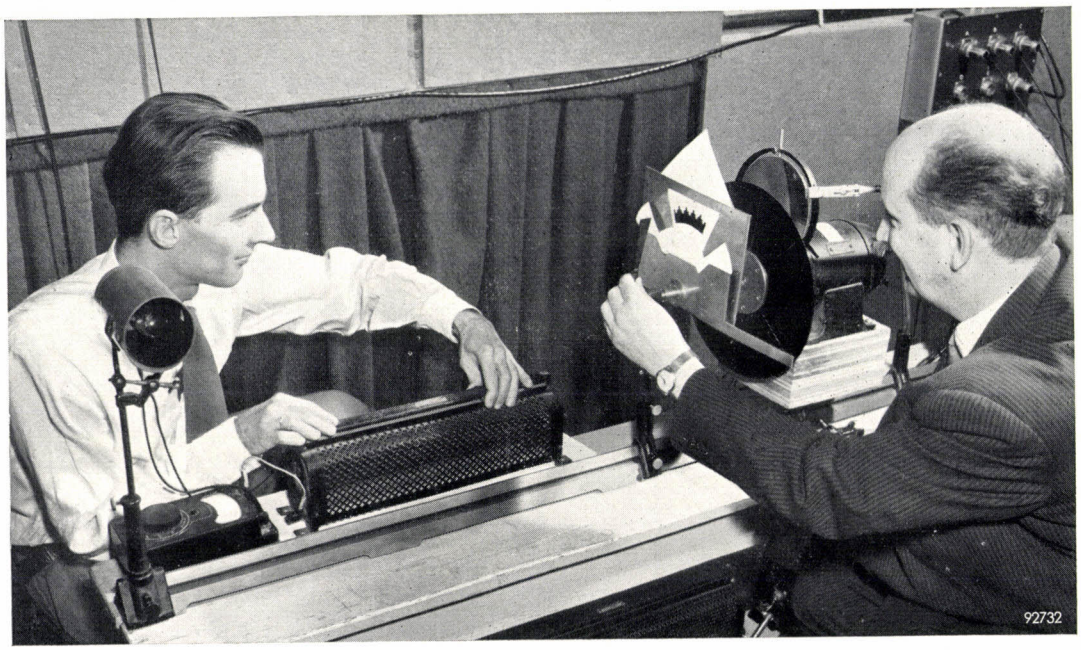
If the arrival rate of barium and barium oxide to the surface never falls below the minimum required to keep the surface completely covered with barium, then the life of the cathode is determined by the amount of activating material present in it and by the rate at which barium is lost by evaporation during operation. The life is therefore dependent on the operating temperature and for an impregnated plug 1 mm thick at 1130 °C it amounts to more than 15 000 hours.

Summary of I, II, III. The L cathode developed in the Philips laboratory at Eindhoven several years ago consists of a porous tungsten body, whose emitting surface is covered by a layer of barium and oxygen atoms. The continuous evaporation of barium during operation of the cathode is compensated by the dispensation of barium from a compound located in a chamber behind the porous tungsten body. The L cathode can be operated continuously at very high current densities, its metallic emitting surface is robust and can be made to narrow dimensional tolerances and with good finish; it has a very long life (several thousands of hours), it recovers easily from poisoning and it is highly resistant to bombardment by ions.

The application of the L cathode in tubes, however, is limited by the relatively long time required for degassing and activation, by the difficulties in hermetically sealing the chamber containing the barium dispensing compound and by its limitations in size.

Investigations in the Philips Laboratories at Irvington have led to the development of two new types of dispenser cathodes, working on the same principle as the L cathode (and therefore having the same general electron emission characteristics) but avoiding its drawbacks. The first type is a pressed cathode, made by compacting a mixture (ratio 9 : 1 by weight) of powdered tungsten-molybdenum alloy and barium-calcium aluminate, and heating the pressed mass. The molybdenum content of the alloy (75% by weight) is essential to lower the evaporation rate of barium and barium oxide to the desired level. The second type is an impregnated cathode, made by impregnating a porous tungsten body with molten barium-calcium aluminate. Suitable dispensation of barium is obtained by using tungsten with a porosity of 17% (the porosity is 27% for the tungsten body of the L cathode and 40% for the alloy in the pressed cathode). This porosity value also permits the application of a special technique for machining the tungsten, so that the impregnated cathode can easily be made in a large variety of shapes and sizes, including very small ones.

Both the pressed and the impregnated cathode have much shorter degassing and activation times than the L cathode and can be manufactured more readily. A pressed cathode of flat shape especially lends itself to mass production, but mass manufacture of the impregnated cathode likewise does not involve any great difficulties, while the latter type offers by far the greatest latitude in design. The calcium content of the aluminate in both types of cathodes enhances the electron emission considerably, by a mechanism which is not yet fully understood. These cathodes can also be exposed to the air and reactivated after restoring the vacuum several times during their life. The electron emission of all three types of dispenser cathodes (L cathode, pressed and impregnated cathodes) makes them suitable for use in tubes which call for high current densities, either continuously or in pulses, for several thousand hours.



ELECTRONIC MUSIC

by H. BADINGS *) and J. W. de BRUYN.

789.983:621.395.625.3:
681.84.087.27

*The term "electronic music" is sometimes used when compositions, written for conventional musical instruments, are performed with the aid of electronic devices such as the Hammond organ, the "Ondes Martenot" or electronic music synthesizers **). In recent years, however, the term has come to mean something quite different, being applied to music whose actual composition is based on the use of electronic aids, especially of magnetic recording. These aids can be so employed as to produce entirely new sounds, of which neither the Hammond organ nor any conventional instrument would be capable.*

One of the leading figures in this new field of electronic music is the Dutch composer Henk Badings. His radiophonic opera "Orestes", in which electronic aids play a substantial part, was awarded the Prix Italia in 1954 and has since received more than 200 performances. In his ballet music "Cain and Abel" — which had its first public performance in May 1956 at The Hague — he makes almost exclusive use of electronic aids. Badings' decision to undertake such an experiment evidently owes much to the fact that, having studied engineering at Delft, he has always been keenly interested in technological problems, particularly in those of an acoustical nature.

The fairly extensive range of electronic instruments required to produce "Cain and Abel" was placed at Badings' disposal by Philips Research Laboratories in Eindhoven. It washere that the score was realized jointly by the composer and the co-author of the present article and recorded for performance on magnetic tape. This article gives a description of the electronic aids employed. By way of illustration to the text, a gramophone record giving a selection of the sounds produced is included in this number of the Review. The record also gives a shortened version of the ballet music itself.

In view of the aversion to dogma and the passion for "development" and innovation that characterize Western civilization in our century, no one will be surprised to hear new notes being struck in contemporary music — literally and figuratively. The element of novelty in this connection has two aspects. There is in modern music, just as in modern poetry and art, a perceptible tendency to break with tradition and to shun classical forms of expression and aesthetic laws. On the other hand — and in this respect music might be likened to architecture, with which it has, in fact, frequently been compared — the work of contemporary composers reveals a direct connection with the most conspicuous element of our civilization — technological development. Several composers are at present making use of modern technical equipment in an endeavour to widen the scope of musical expression.

*) Ir. H. Badings, Bilthoven, Netherlands.

) For the Hammond organ and similar electronic musical instruments, see: A. Douglas, The electronic musical instrument manual, Pitman & Sons, London 1957 (3rd ed.). For music synthesis, see H. F. Olson and H. Belar, Electronic music synthesizer, J. Acoust. Soc. Amer. **27, 595-612, 1955.



Fig. 1. Part of the equipment employed in the Philips Research Laboratories for making the electronic ballet music "Cain and Abel". On the table are some of the seven magnetic tape recorders used (sometimes five at a time). Also to be seen are a number of amplifiers, a mixing desk, filters, metronomes, various electronic sound-sources, loudspeakers, etc.

This is not the place to go into the evolution of the latter type of music. Those who wish to know more about the subject can find information in various books and published articles [1]). It will be enough to note here that the different trends of composition that can at present be distinguished all agree on one essential point: they invariably employ *electronic* means, and in particular *magnetic recording*, for transforming the sounds and moulding them into something entirely new and unique. All these trends will be referred to by the collective term "Electronic Music" [2]). The differences of view between the protagonists of the various trends, some of whom, curiously enough, have brought a new dogma into being, we can only touch upon in passing. Schaeffer in France wants the composer's material to be derived from "concrete" sounds, as for example a steam whistle, a clap of thunder, a dripping gutter, and so on, and he declares all traditional musical conventions taboo ("Musique Concrète"). The "Studio des Kölner Rundfunks", where Eimert and Stockhausen are at work, wants to admit only synthetic sounds produced electronically; this school of thought also seeks to avoid traditional musical elements such as melody, rhythm and harmony, and favours the twelve-note technique introduced by Schönberg in about 1924. Among the composers active in this field in other countries may be mentioned Varèse, Luening and Ussachewsky in the United States, Maderna and Berio in Italy, and in the Netherlands Badings (radiophonic opera "Orestes", Prix Italia 1954; ballet music "Cain and Abel"; film music "The Flying Dutchman") and De Leeuw (radiophonic oratorium "Job", Prix Italia 1956). The less dogmatic attitude of these composers appears from the fact that they sometimes, as in the radiophonic works mentioned, use electronic means in combination with, or only as complementary to, traditional musical instruments.

In this article we shall be chiefly concerned with the technical aspects of Electronic Music. It is evident that the wider the range of electronic resources on which the composer can draw, the richer can be the creative development of this music. For the composition (and "performance") of the ballet music "Cain and Abel", which had its première in 1956, the composer drew on the equipment and cooperation of the Acoustics Group of the Philips Research Laboratories in Eindhoven. The principal apparatus (see *fig. 1*) employed in "Cain and Abel" are described in the following pages.

Examples of the sounds produced are given on the gramophone record appended to this article.

[1]) P. Schaeffer, A la recherche d'une musique concrète, Ed. du Seuil, Paris 1952.
W. Meyer-Eppler, Elektrische Klangerzeugung, Dümmlers Verlag, Bonn 1949.
H. Badings and A. Brandon, Concrete, electronische en radiofonische muziek, Radio Electronica **4**, 144-147 and 152, 1956 (No. 3). (In Dutch.)
F. K. Prieberg, Musik des technischen Zeitalters, Atlantis Verlag, Freiburg 1956.

[2]) The reader should be reminded that the term has been given various other definitions.

The index figures [..] in the text refer to these examples. The reader is recommended to read the article first and then to listen to the record while referring to the condensed explanatory notes given at the end of the article.

Technical aids to Electronic Music

Electronic engineering enters into the creation of Electronic Music in three ways: it provides new sources of sound, it makes it possible to manipulate and transform sounds, and finally it governs the reproduction of the music. For our purposes it will be convenient to begin at the end, that is to discuss the reproduction first.

Reproduction

The term "electronic reproduction" implies the re-creation via amplifiers and loudspeakers of sound originally created at another place or at another time. The appreciation of every type of "reproduction" is primarily determined by the criterion of its fidelity to the original. The familiar shortcomings of electronic reproduction in this respect (noise, distortion and the hole-in-the-wall effect[3])) have been eliminated to such an extent by improvements, for example, in electronic tubes and in recording methods (magnetic recording) and by the introduction of stereophonic techniques, that electronic reproduction can now be made almost indistinguishable from the original. Now that the odium of being a mere substitute need no longer attach to electronic reproduction via loudspeakers, more stress can be laid on the fact that such reproduction offers possibilities which are denied to the original (in this sense it would really be more appropriate to speak of electronic "production" than of "reproduction").

One such possibility resides in the unprecedented freedom with which the position of the sound source or sources can be arranged in a given space. This has been exploited — to give one example — by disposing loudspeakers around an auditorium and by feeding these loudspeakers with electronically delayed music in such a way as to improve the acoustics of the auditorium. This has been done both for "live" music[4]) and for the reproduction of recorded music (stereo reverberation). The composer might go a step further along these lines by not only arranging for different parts of the overall tonal pattern of his score to reach the listeners from different directions and distances — a generalization on the idea of the trumpeter behind the wings in "Fidelio" — but by also giving these parts a different spatial character. It is possible in this way to produce all transitions between the clear and dry directional sound of a trumpet in the open air and the diffuse, non-directional sound of choral music in a church. Something of the kind was in fact done in the performances of "Cain and Abel" at The Hague and Hannover.

Electronic reproduction offers another possibility of the utmost practical importance: in the electronic circuits of the apparatus, the music — or its component sound signals — is available in a form in which its strength (volume) can be controlled by the extremely simple expedient of turning a potentiometer knob. Volume control, and with it the mixing of different parts of a tonal pattern in any desired proportions, is a long-established practice in broadcasting studios. In the early days of broadcasting there was indeed some concern in the artistic world about the power given to the man at the mixing desk to alter the balance of a piece of music. If the mixing engineer has had a good musical training there need be no fear that he will misuse his powers, but from concern about a possible misuse it is only a small step to realize that in the hands of the composer these powers can be turned to useful account. With the volume control he can introduce dynamic figures in notes or combinations of notes where this would otherwise be impossible, as for example in a dying piano chord [1]. By a rapid movement of the volume control, he can deprive a note struck on a percussion instrument of its "attack", and in so doing give a pianoforte passage, for example, a quite unique character [2]. Moreover, there is now no longer a fixed relation between timbre and intensity, as was hitherto the case with many instruments: a whispered passage of a song can be made to sound above an orchestral tutti, although it still retains its "whispered" characteristics.

In the ballet music "Cain and Abel" repeated use is made of a dynamic figure in which the sound intensity is given a bell-shaped variation in time, i.e. it swells rapidly and immediately dies away again. This figure is produced by automatically varying the volume with a device referred to as a tone gate, the circuit of which is given in *fig. 2*; the duration of the bell-shaped variation it produces can be selected in five steps.

The electronic treatment and transformation of sounds

We have seen that in electronic reproduction the sound is at hand in a form in which it can easily be controlled in volume. When the sound is repro-

[3]) See e.g. R. Vermeulen, A comparison between reproduced and "live" music, Philips tech. Rev. **17**, 171-177, 1955/56.
[4]) R. Vermeulen, Stereo reverberation, Philips tech. Rev. **17**, 258-266, 1955/56.

duced via the medium of magnetic tape, an even greater measure of control is possible. Originally intended as a means of recording speech and music, for playback as and when required; magnetic tape provides the conductor or composer with remarkable possibilities: they now have the music literally at hand from second to second and can do with it whatever they wish. A good illustration and one that is of considerable practical importance, is the ability to cut out of the tape an unsuccessful passage of a recorded work and to insert in its place a better executed version. In recent years this

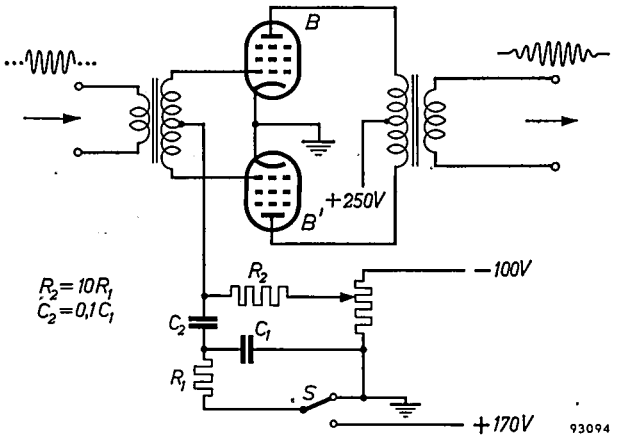


Fig. 2. Basic circuit diagram of the "tone gate". A signal, e.g. an alternating voltage of constant amplitude (constant sound intensity) is applied to the input of a push-pull amplifier with tubes B and B'. In the position of switch S as shown, the tubes are biased below cut-off by a voltage of about −90 V on the grid and no signal appears at the output. When switch S is thrown over, the grid bias, rapidly at first, goes more positive and rises above cut-off potential, whereupon (depending upon the charging of capacitors C_1 and C_2) it returns again to −90 V. During that time the amplifier gain rises gradually from zero to a certain value from which it at once gradually returns to zero, so that the input signal arrives at the output with its amplitude modulated in the shape of a bell. The width of this tone pulse (i.e. the duration of the tone or note) can be regulated in five steps by switching over to other values of capacitors C_1 and C_2. After returning the switch to its original position (whereupon the capacitors discharge very rapidly via diodes, not shown) the same process can be repeated.

practice (editing) has brought about a veritable revolution in the gramophone record industry [5]). But the composer can also exploit the same procedure in order to create new musical sounds: for instance, with each note of a percussion instrument he can remove the piece of tape on which the attack is recorded, thereby producing an effect which, though of the same nature as that described in the last paragraph, is yet quite different [3]. A very striking transformation is obtained by playing a piece of tape at a speed different from the recording speed [4]. The pitch, of course, changes proportionately, but at the same time an unusual tonal pattern is produced owing to the fact that the relative strength of all harmonics remains exactly the same, which is not generally the case when playing different notes on an instrument. Accelerating the tape also makes it possible to play extremely fast runs with a perfection that is quite beyond the reach of a flesh and blood instrumentalist. Continuously varying the tape speed results in a glissando which, when applied for example to notes played on a pianoforte, produces a most remarkable effect [5]. (This was done by using a tape recorder with a variable frequency power supply for the motor.) The tape can also be played in reverse: the notes of a pianoforte passage then swell up successively, each ending on the "attack" [6]. Another possibility with magnetic recording is to conserve the sound for a very short time on, for example, a rotating magnetic wheel, the sound being taken from the wheel by a number of playback heads and added in variable intensity to the signal being recorded or reproduced (*fig. 3a*). In this way a reverberation [7] is introduced, on the same principle as the method of stereo reverberation employed for improving the acoustics of an auditorium [4]). If the signal taken up by a playback head is fed back to the recording head with a longer delay and greater loop gain (fig. 3*b*), the result is a sound something like the familiar "motorboating" of an oscillator [8].

Magnetic tape is undoubtedly the most important aid to the creation of Electronic Music, not only because of the wide scope it offers for the transformation of sounds, but also because it was with this medium that it first became possible to register individual sounds, to transform them and then to put them together to form a unified whole — in other words, to "compose" them in the musical sense. We shall return presently to this procedure of composition and discuss its consequences from the artistic point of view.

The manipulation of magnetic tape is not the only method of transforming sounds. An obvious method is to use electrical networks, with which variable linear distortions can be effected ("shaping" of the frequency characteristic). A special case in point are *electrical filters* with passbands characterized by sharp or gradual cut-off, beyond which the sound is attenuated or even suppressed [9]. By this means a particular narrow range of frequencies in each note can be given a dominant significance (creation of formants).

Another sound-transforming device which was also used in the production of "Cain and Abel",

[5]) See e.g. J. L. Ooms, The recording and production of gramophone records, Philips tech. Rev. **17**, 101-109, 1955/56.

is the *modulator*, in the form familiar for example, in carrier-wave telephony [6]). This modulator has two equivalent pairs of input terminals; if two musical tones are applied to these terminals, both tones appear at the output together with all their combination tones (intermodulation products). Of these the difference tone (beat frequency) is the most prominent. If the one input signal is a musical chord and the other a purely sinusoidal vibration, the chord will appear at the output accompanied by a kind of shadow chord which is a constant frequency lower [10].

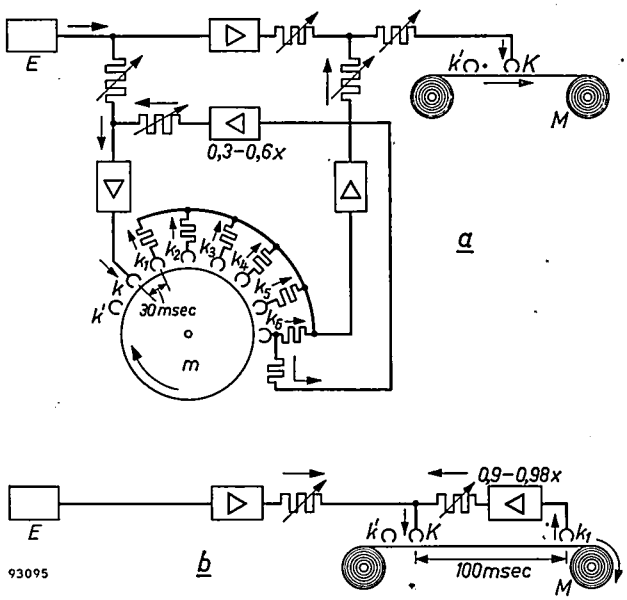


Fig. 3. Arrangements for introducing reverberation (*a*) and excessive feedback (*b*). *E* sound-source (this may be an amplifier if the sound-source is not electronic itself); *M* magnetic tape on which the result of the treatment is registered by recording-head *K*.
a) The signal is also registered by recording-head *k* on a magnetic layer around the rim of a drum *m* with erasing head *k'*; the signal on this layer is then picked up successively by a number of playback heads k_1-k_6, situated at progressively increasing distances from *k*. These signal components, delayed by 30 up to 180 milliseconds, and which are not separately distinguishable but together create the impression of reverberation, are fed, with the original signal, to *K*. In general, reverberation times much longer than 180 milliseconds are needed; to avoid using a very large number of heads for this purpose, the signal picked up by the last playback head k_6 is passed back to recording head *k* (loop gain 0.3 to 0.6×), resulting in six weaker components with delays up to 360 msec, followed by six more even weaker components with delays up to 540 msec, and so on.
b) A feedback loop is formed on the magnetic tape itself between playback head k_1 and recording head *K*, with a delay time of at least 100 msec, and an appreciably larger loop gain (0.9 to 0.98×); in this case the fed-back signal can be heard separately and sounds rather like the "motor-boating" of an oscillator.
With the reverberation equipment in (*a*) one feedback head is not enough, since the overall frequency response is not completely flat and hence the signal cannot be fed back too many times.

[6]) See e.g. F. A. de Groot and P. J. den Haan, Modulators for carrier-telephony, Philips tech. Rev. 7, 83-91, 1942.

Electronic sound-sources

By the manipulation of magnetic tape quite new "sounds" (or rather, vibrations) are produced which can be made audible via a loudspeaker; they are effectively "original" sounds although they were never picked up by a microphone. Once we are familiar with this idea, it soon becomes obvious that we can dispense entirely with the primary sound that can be picked up by a microphone and, instead, apply directly to the loudspeaker such vibrations as can be produced by various electronic devices. A number of electronic sound-sources were used in the composition of "Cain and Abel", some of which have long been known and widely used, although chiefly for other purposes.

In this sense the Hammond organ and kindred instruments are also electronic sound-sources and the composer of "Electronic Music" could, if he wished, use them just as well as conventional musical instruments. When seeking for new tonal possibilities, with the aid (*inter alia*) of sound transformations, there is little point, however, in starting with a rather intricate instrument on which a great deal of effort has already been spent with the object — again by electronic means — of producing or imitating certain timbres of musical sounds. The sound-sources which we have used are therefore of a more elementary and less specialized nature.

From the physical point of view the simplest electronic sound-source conceivable is the *sine-wave generator* (commonly known as an audio signal generator). A loudspeaker connected to such an apparatus delivers an almost purely sinusoidal sound pressure, that is to say a tone practically devoid of harmonics, which strikes the ear as strange and ethereal [11]. The pitch of the tone can be continuously varied by simply turning the frequency control knob. By this method glissandi [12] can be produced at will, but its greatest virtue is that it offers a simple means of obtaining with great accuracy notes with various intervals, i.e. any desired scale. One can therefore deliberately depart from the narrow paths of the conventional equal temperament scale, which divides the octave into twelve equal intervals. In "Cain and Abel" repeated use is made of "pure" (harmonic) intervals, that is to say intervals exactly equal to the whole-numbered frequency ratios of the harmonic series. Although some of these intervals, used melodically, sound somewhat unusual, they are soon accepted by the ear as "good". We mention specially the use of intervals with ratios in which the numbers 7 and 11 appear, which have no counterpart in European music: they have hitherto been rejected since, although they blend perfectly in a chord of harmonically tuned notes, they are strongly dissonant with

certain notes in the equal temperament scale (e.g. the seventh). (For this reason a pianoforte is designed so as to suppress as far as possible the "dissonant" seventh harmonic of the fundamental of each string.)

To achieve harmonic tuning — as, indeed, for many other purposes of checking — we made use of a cathode-ray oscilloscope. A harmonic interval produces on the oscilloscope screen a typical stationary



Fig. 4. The composer playing a multivibrator.

Lissajous figure. This method does away with all the intonation difficulties that arise if an instrumentalist is asked to play, by ear, unusual intervals with the traditional continuously tunable sound-sources (the singing-voice, violin, etc.).

The counterpart of the sine-wave generator is another familiar electronic device, which is widely used, for example, in counter circuits — the *multivibrator*. This produces a sound which contains all harmonics of the fundamental up to the audio limit, the successive harmonics differing in intensity according to some slowly varying function. (The resulting signal, which may have a square or a sawtooth waveform, can also, of course, be modified by passing it through an electrical filter or shaper circuit.) Two versions of this sound-source were used in the composition of "Cain and Abel". The first, which has been given the name "baritone clavier", supplies a signal having an almost square waveform, the fundamental of which can be adjusted in pitch to a series of discrete frequencies by means of push-buttons [13]. With the second instrument the pitch of the fundamental can be varied continuously (making glissandi possible); in the low register the vibration approaches a square waveform, but in the high register it gradually assumes a sawtooth form [14].

A photograph and the circuit diagram of this multivibrator can be seen in *fig. 4* and *5*.

For various parts of "Cain and Abel" use was also made of a *noise generator*. As a sound-source this device differs from all other known sources in that the sound it produces has a continuous instead of a line spectrum. Noise having a constant energy distribution over the entire spectrum of audible frequencies ("white noise") is musically not particularly interesting since there is nothing that can be varied except its intensity; nevertheless, with the aid of filters, it is possible to impose on the noise a somewhat vague impression of pitch. To do this we employed filters passing a frequency band of one or several octaves [15], or a third [16], these bands being capable of displacement in certain fixed steps.

In the electronic sound-sources mentioned above, the frequencies of the vibrations, with all their components, are determined by *electrical* elements. In another, rather hydrid, group of electronic sound-sources the frequencies (pitch and tone colour) are determined by *mechanical* means, although the vibrations only become acoustically effective through the medium of a loudspeaker. To this group belong two kinds of "electronic drums" and an "electronic clavichord", which were used for "Cain and Abel". The first drum was essentially just a large condenser microphone, the vibrations

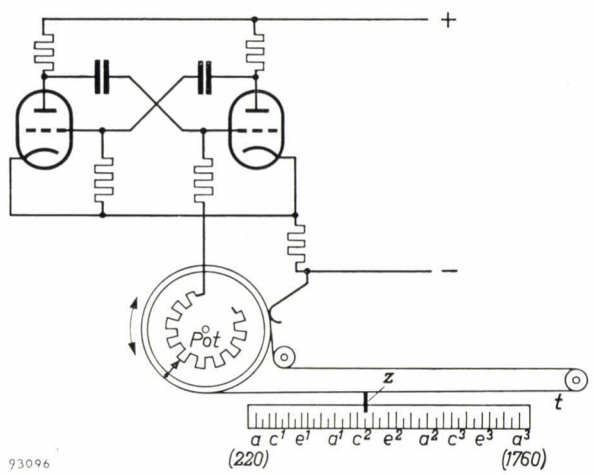


Fig. 5. Circuit diagram of a multivibrator with continuously variable pitch. The pitch is controlled by a potentiometer *Pot* whose sliding arm is fixed to a wheel; a cord and pulley system enables the musician to turn the wheel by moving a clip z along a scale calibrated in notes from a to a^3. By moving z rapidly to and fro a natural vibrato can be produced.



Fig. 6. The electronic clavichord.

being produced by taps or drum rolls on the diaphragm. The second consisted of a steel sheet whose vibrations were picked up at a specific position by a piezo-crystal; according to the part struck and the manner of damping the vibrations, a considerable variety of sounds could be produced [17]. A photograph of the "electronic clavichord" [18] is shown in *fig. 6* and a description of the device is given in the caption to *fig. 7*, which shows the circuit diagram.

The last to be mentioned of our range of electronic sound-sources is a particularly remarkable instrument which we call an *optical siren* [7]). In this device the pitch is determined mechanically, namely by the speed of a motor-driven disc, but the tone colour is produced by optical means, a pattern cut out in a sheet of paper being successively scanned by slits in the revolving disc placed in a beam of light, and the light-variations so produced being converted into electrical variations by a photo-electric cell (*fig. 8*). "Tone-colour melodies" created with this instrument are to be heard in several parts of "Cain and Abel" [19].

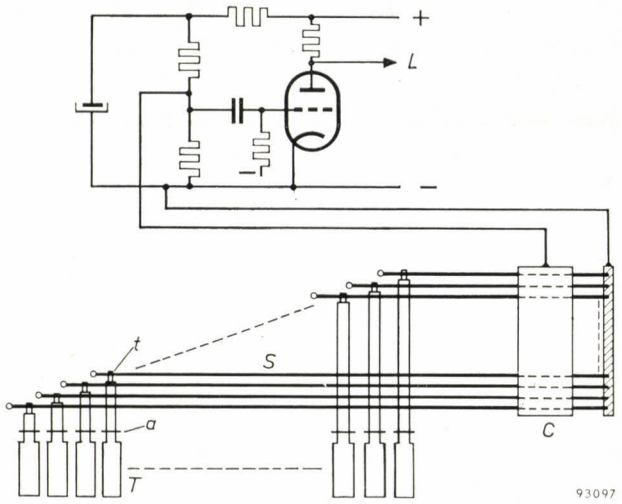


Fig. 7. Schematic representation of the electronic clavichord. The parallel single wires *S* of this instrument are made to vibrate in the same way as in a normal clavichord, i.e. with key-operated tangents (*t*; *T* are the keys and *a* their pivots) which, after being struck, also act as the endpoints of the respective strings. The strings as a whole constitute one electrode of a capacitor, the stiff plate *C* above the strings being the other electrode. This capacitor is incorporated in an amplifier circuit in exactly the same way as a condenser microphone. A loudspeaker can be connected either directly to the amplifier at *L*, or via various electronic devices.

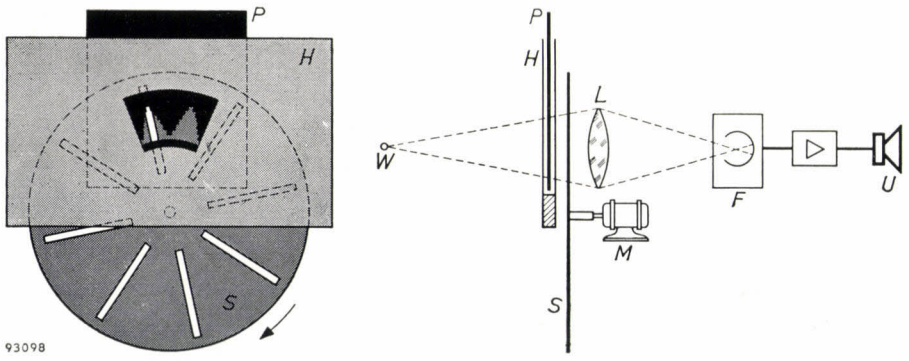


Fig. 8. Schematic representation of the optical siren (see photo on title page). A beam of light emitted by light-source *W* is concentrated on a photomultiplier tube *F* by a lens *L*. Situated in the beam is a holder *H* containing a sheet of paper *P* in which a waveform has been cut out, and behind which rotates a disc *S* driven by a motor *M*. As the disc rotates the pattern is successively scanned by narrow radial slits in the disc, and the light passing through the slit varies in accordance with the ordinate of the pattern. The photoelectric current in *F* varies in the same rhythm and is passed via an amplifier to the loudspeaker *U*.

A natural vibrato effect can be introduced by holding the sheet of paper in the hand instead of fixing it in the holder.

[7]) J. F. Schouten, Synthetic sound, Philips tech. Rev. **4**, 167-173, 1939.

The composing and "performance" of electronic music

Outwardly the most striking difference between an "electronic" and a traditional composition is that the former cannot be performed by musicians in the presence of the public. The process of transforming sounds, for example by cutting and splicing, retarding, accelerating or reversing magnetic tape, excludes this possibility at once and leads to quite a different procedure. The composite sounds which the composer indicates at a given moment in his score — for example, a melody from the signal generator, an accompanying pianoforte figure, on which is to be superimposed rhythmical figures from the clavichord with accenting by an electronic drum — are each produced separately by a musically trained engineer and transformed on instructions given by the composer. The result is a number of "acoustic layers" (in this instance four), each recorded on a separate magnetic tape [20]. Tuning control by means of fixed frequencies, and accurate time measurement ensure that the acoustic layers are properly matched and synchronized (in the composition of "Cain and Abel" the tuning was checked with the aid, among other things, of a set of electrically driven tuning forks which produce tones from c^1 to c^2). The layers are then put together in the correct dynamic relationship either all at once or, if preferred, one by one [21], until at last a single magnetic tape is obtained on which the entire work is recorded. This tape is now ready to be played to the public.

A frequently heard misconception is that this procedure must lead to a rigid, mechanical tonal pattern in which there is no room left for artistic interpretation. It cannot be denied that direct contact between performer and audience is lost owing to the prior performance of the work and its reproduction via magnetic tape, although the same applies to broadcast music and to music played on gramophone records. It is also true that, in the case of "Cain and Abel", for example, the work was largely performed for the first time by the composer himself — partly because there are as yet so few musicians who can "play" electronic apparatus, and partly because the act of composing in this new realm of music sometimes runs parallel with the act of performance [8]). Nevertheless, there is certainly room for interpretative re-creation, as testified by the fact alone that the composers of Electronic Music do not identify their work with a magnetic tape but set down the musical thought they wish to express in the notation of a score (*fig. 9*), even though, in doing so, they must necessarily use many new signs, often inventing them themselves. The scope for interpretative re-creation appears in the fact that in making as well as in putting together the acoustic layers indicated in the score there are innumerable details, for example the question of relative sound intensities, that must be decided by musical feeling. This has already been made plain by the results of making versions of the radiophonic opera "Orestes" in other languages: the individual performances (in Dutch, German and English) reveal very striking differences of interpretation in the electronic sounds, although in none of them were the composer's intentions misrepresented.

The composer of Electronic Music need scarcely offer a justification of his work. From the psychological point of view he can regard his experiments in this unexplored territory as sharing in that passion for innovation and development which, as we remarked at the beginning, is characteristic of Western civilization. He can supply their musical motive by pointing to the charm or the fascination of the new sounds, and in particular:

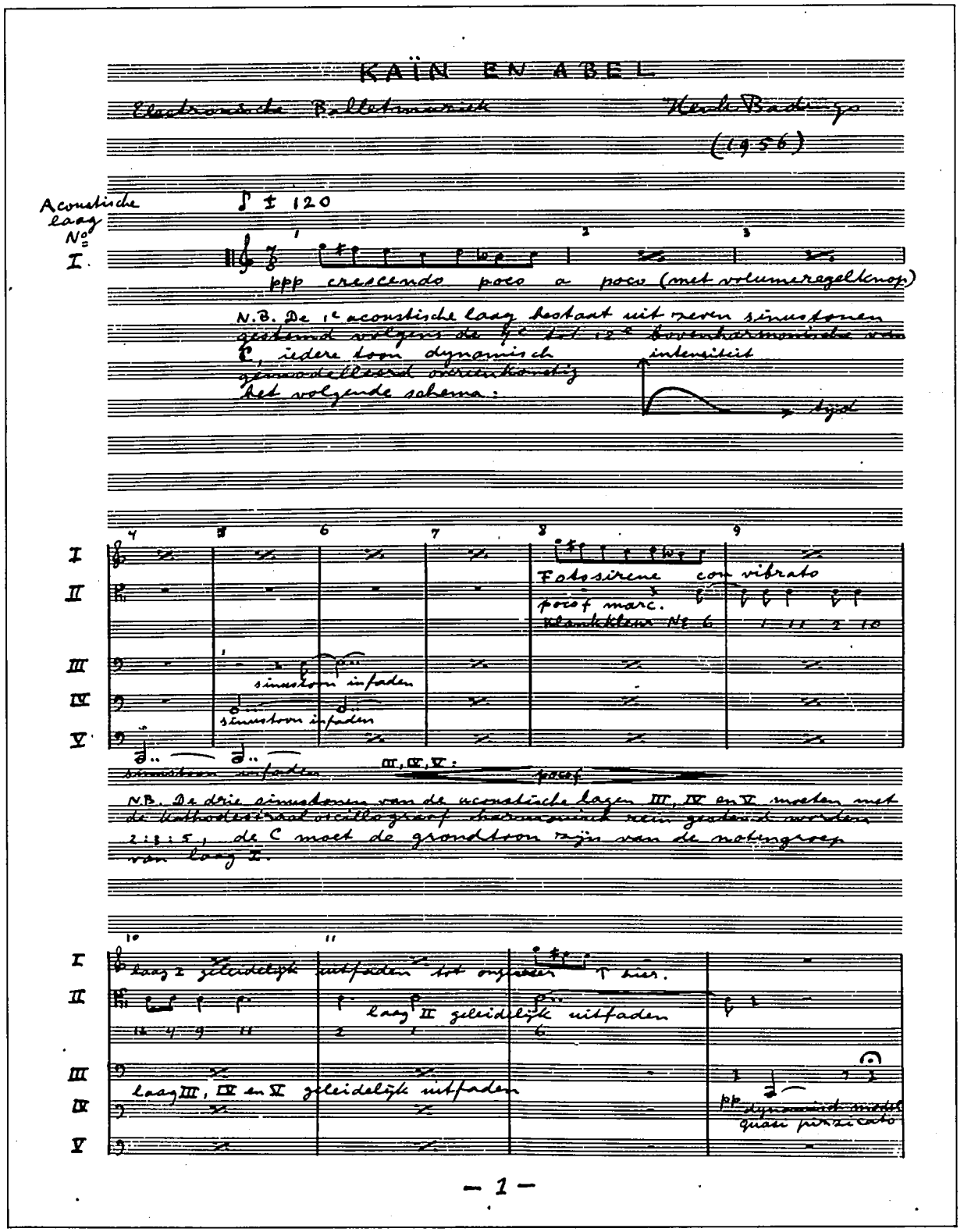
the timbres with their rich variations;
the harmonic tuning;
the wide scope in the dynamics of the music;
the tempi and the brilliance of the figures, not subject to human limitations;
complete freedom to combine independent rhythms (whereby the metrical accents on the average can even cancel each other out, producing a curious, hovering movement [22]).

Having, therefore, regard to the common definition of music — as the art of conveying artistic emotions from man to man by the medium of air vibrations — there can never be any doubt in the mind of the composer of electronic music about the right of his brain-child to exist, but he must leave the decision to the listener — of today or perhaps of tomorrow.

Explanatory notes to the acoustical illustrations

Side 1 of the gramophone record appended to this article contains examples of individual sounds touched upon in the text; each example is introduced by a *morse-code signal*, by which it can be identified in these explanatory notes. The index figures [..] refer back to the text. Side 2 contains a much abridged version of the ballet music "Cain and Abel"; it was put

[8]) As a matter of interest, the process of composing and making the magnetic tape for "Cain and Abel" was completed within one month. In general, however, the realization of an electronic score is a rather time-consuming process, and roughly speaking requires just as many "man hours" as it does to rehearse with an orchestra a traditional concert piece of comparable length and complexity.



92797

Fig. 9. Page 1 of the score of the electronic ballet music "Cain and Abel". It can be seen that five "acoustic layers" are used and that the composer of this kind of music is at present obliged to add a lot of written explanation to his score in order to specify his intentions. For the optical siren ("Fotosirene" in the score) twelve vibration patterns (each producing different tone colours — "klankkleur" in the score) were cut out beforehand and numbered, so that it was only necessary to indicate in the score the numbers and the pitch (r.p.m. of the disc).

In other parts of the score other methods of notation are used, including a kind of tablature (a finger-position notation as formerly used for the lute) and a graph in which the required vibration frequencies are directly plotted as a function of time.

together by the composer from the original tape (playing time about 20 minutes) so as to offer in a playing time of 8 minutes a large number of striking passages without doing too much violence to the musical form of the whole.

The examples on side 1 fall into three groups:

I. Illustration of the effect of electronic sound treatments, each applied to the same traditional musical sound, viz. a pianoforte figure, recorded on magnetic tape.

II. Illustration of the sound of several of the electronic sound-sources discussed in the text; some are the original sounds direct from the source, others are treated as under I.

III. Demonstration of the combining of acoustic layers. A passage from "Cain and Abel" is chosen which occurs on side 2 of the record; to make the demonstraton clear, however, the layers are put together anew and some treatments prescribed in the score (including the addition of reverberation) are omitted.

Group I (Various electronic treatments)

First the pianoforte figure itself is heard, consisting of a chord, a run, and the same chord an octave higher.
Then:

—• Tape retarded (speed halved) [4].
Acceleration (tape speed doubled and tape played twice) [4].
Acceleration (tape speed quadrupled and tape played three times) [4].
Glissando of the chord over one complete octave [5].

—•• Dynamic figure introduced in chord by varying the volume control [1].
Attack of each note in the run suppressed by means of the volume control [2].
Attack of the same notes removed by cutting out the pieces of tape at which each attack occurs [3].

—••• Reversal of time by playing the tape in reverse [6].

—•••• Excessive feedback applied to the chord (played twice) [8].
Reverberation added to complete piano figure [7].

Group II (Various electronic sound-sources)

——• Sine-wave generator: staccato notes [11], glissando [12], legato notes [11].

——•• Optical siren: "tone-colour melody" [19].

——••• Multivibrator ("baritone clavier") [13].

——•••• Multivibrator (continuously variable) [14].

——••••• Steel sheet, beaten and damped in different ways [17], also in part electronically treated by the introduction of dynamic figures and by accelerated playback (cf. passage near the end of side 2).
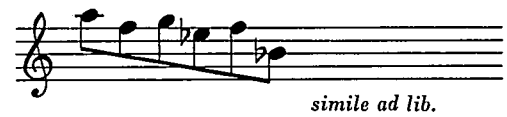
——•••••• Noise generator:
Bandwidth 1 octave (successively 8-4, 4-2, 2-1 and 1-0.5 kc/s) [15].
Bandwidth 1 third (mean pitch varying in steps from 100 to 1600 c/s) [16].

——••••••• Electronic clavichord [18]:
The same figure as the pianoforte figure in Group I.
A different figure; magnetic tape played with continuously increasing speed.
Intermodulation of the last figure and a sinusoidal tone of rising frequency.

Group III (Combination of acoustic layers)
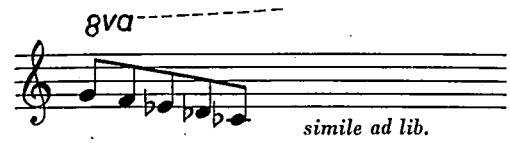
— A complete passage, composed of five layers.

—• First layer:
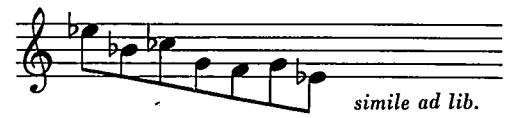


obtained from pianoforte figure played at 224 quavers per minute (♪ = 224); tape played at double speed in reverse [20].

—•• Second layer:
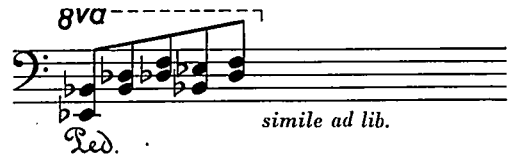


obtained from pianoforte figure produced by depressing the keys of the chord on the keyboard and scraping a hard object over the strings involved, ♪ = 200; tape played at double speed.

—••• Third layer:



obtained from pianoforte figure played at ♪ = 65; tape played at double speed.

—•••• Fourth layer:



pianoforte figure played at ♪ = 105; normal tape speed.

—••••• Fifth layer: sound consisting of notes struck on tubes, similar to the chime of a bell in E flat, followed by the clavichord figure [18] with intermodulation given at the end of Group II, but with tape speed increasing still further.

——• Combination of third and fourth layers [21] [22].

——•• Combination of third, fourth and first layers.

——••• Combination of third, fourth, first and second layers.

———• All five layers combined, same as heard at the beginning of this group.

Summary. In contemporary music various trends of composition have evolved which may be described under the general heading of "Electronic Music". A characteristic common to them all is the use of magnetic recording and other electronic aids for the treatment and transformation of sounds into something entirely new. The composers "raw material" may be produced by traditional musical instruments, or derived from "natural" sources or created by electronic means. In this article the electronic instruments are described which the composer, Henk Badings, used in the Philips Research Laboratories, Eindhoven, for creating his electronic ballet music "Cain and Abel". The authors deal in some detail with a number of possible treatments, such as retarding, accelerating and

N.V. PHILIPS' GLOEILAMPENFABRIEKEN

reversing, making dynamic figures, filtering, intermodulating, producing feedback effects and adding reverberation. They then go on to describe some electronic sound-sources, including the sine-wave generator, multivibrators, noise generators, the "optical siren" and an electronic clavichord. The composer sees the musical significance of using such devices in the wide scope they offer, e.g. for adding innumerable new timbres to traditional musical instruments, for working freely with scales differing from the conventional equal temperament scale, for producing faster and more brilliant figures than would otherwise be humanly possible and for employing extremely complicated rhythms that could scarcely be maintained by even the most skilful ensemble.

The various possibilities are illustrated by acoustic examples given on a gramophone record appended to this article, and which also offers an abridged version of "Cain and Abel".



AN OPEN-AIR LABORATORY FOR ROAD LIGHTING

628.971.6.006.2

At the new Philips factory at Turnhout (Belgium), which manufactures sodium and mercury-vapour lamps, there is an open-air lighting laboratory, now in use for more than six months. As can be seen from *fig. 1*, the open-air laboratory consists essentially of a broad road (320 m long and 14 m wide) provided with ten lighting masts mounted on trolleys. The arms of these masts, on which the fittings are mounted, can be adjusted in height by means of electrical hoist mechanisms contained in the trolleys. The adjusted height can be read on a dial; the maximum height is 12.5 m above the ground, and the minimum height, for the convenience of changing the fittings, is 1 m (see *fig. 2*).

The laboratory serves a fourfold purpose, being employed for fundamental research in the field of road lighting, for judging the performance of new fittings and lamps, for providing assistance in the design of road-lighting installations and for demonstrations.

Fundamental problems, such as the nuisance of glare on lighted roads and the question of what uniformity of surface luminance (brightness) is required for good visibility, can only be partially studied on scale models. At a certain stage a test under actual conditions will be necessary, and that is where the open-air laboratory proves its value.

The open-air laboratory provides the facilities for checking the behaviour of new designs of lamps



Fig. 1. Daylight view of the Philips open-air laboratory for road lighting.

and fittings against the expectations of photometric observations indoors. This in turn offers a firmer and more reliable basis for making decisions on putting a new product into production.

In the design of road lighting layouts it has been found that calculations alone are not enough to provide a complete picture of the merits of a given installation. If this applies to a situation where the road surface is dry, it is all the more applicable to wet road conditions, which must inevitably be taken into consideration. When it is a question, not of standard layouts for streets or roads of no special character, but of lighting important arterial roads or places where the traffic situation is complicated, it is then particularly desirable to be able to test the efficiency of a proposed installation under conditions which closely approach those which will obtain in practice. For this reason the introduction of a major lighting system is often preceded by a trial layout along a short stretch of the actual road; in some cases several such stretches may be fitted out with trial installations having different types of lamps and fittings. This is a costly procedure and clearly not very efficient; it offers no opportunity of making modifications at short notice in, for example, the positions of the lighting poles or in the height of the fittings, in order thus to arrive experimentally at the best solution of the problem. In the open-air laboratory, however, all the necessary facilities are ready to hand.

The open-air laboratory is eminently suitable for demonstrating to interested parties the advantages and disadvantages of different lighting installations. In such comparative demonstrations it must be possible to alter rapidly the spacing and height of the light-sources and also to change the lamps and fittings smoothly and promptly whenever necessary. Having regard to the fact that many of the gas-discharge lamps used for street-lighting need a relatively long time to come up to operating temperature, each trolley contains a large light-tight compartment in which fittings, with lamps already burning, are mounted ready for use on sliding racks (*fig. 3*). A single example will serve to illustrate the scope for demonstrations offered by the many facilities of the open-air laboratory. A trial installation was recently set up along a certain section of a large new motor-road in the Netherlands. After completion, this installation was found to fall far short of expectations. When our Lighting Laboratory was consulted on the matter, a demonstration was arranged at the open-air laboratory where, in the space of two hours, four lighting systems were shown (including that already tested and rejected).

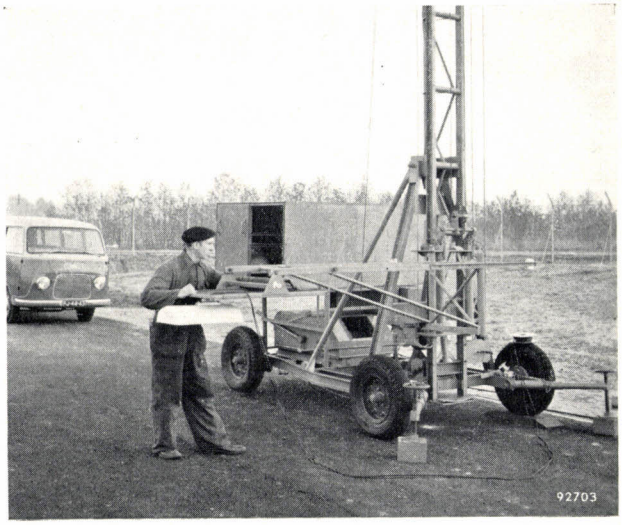


Fig. 2. Changing a fitting in one of the masts. The arm is brought down to its lowest position on the mast. The fitting is mounted on a frame which is slid on to the arm. The frames also fit in the slots at the sides of a compartment in the trolley; the compartment is made light-tight by a sliding panel (see also fig. 3). A van equipped with measuring instruments can be seen on the left (see also fig. 5).

We shall now touch briefly on some details of the equipment of the open-air laboratory and on the measurements that can be carried out there.

The experimental road has a dark (asphalt) surface for 200 metres of its length, the remaining 120 metres having a light surface (a porphyry granite gravel). A general visual impression can thus be obtained of the effect of a lighting system on both dark and light road surfaces. Luminance measurements (which are of course of the utmost importance for appraising a given

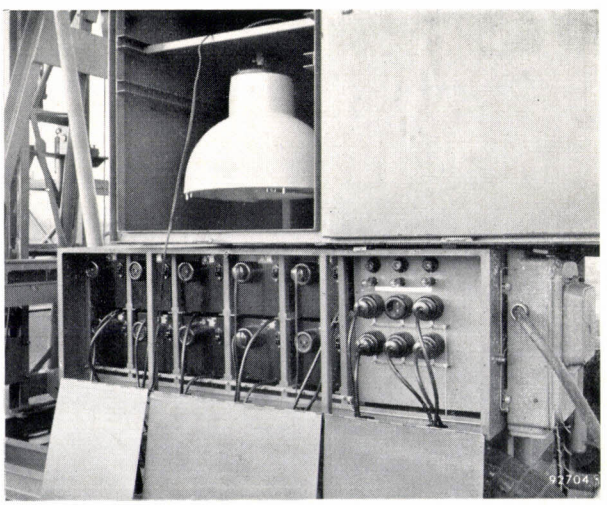


Fig. 3. A fitting mounted on its frame in the light-tight compartment on the trolley, where several fittings with lamps already burning can be stored. When changing a fitting, then, no time is lost waiting for the lamp to come up to temperature. Below the compartment can be seen the various ballasts needed for the lamps.

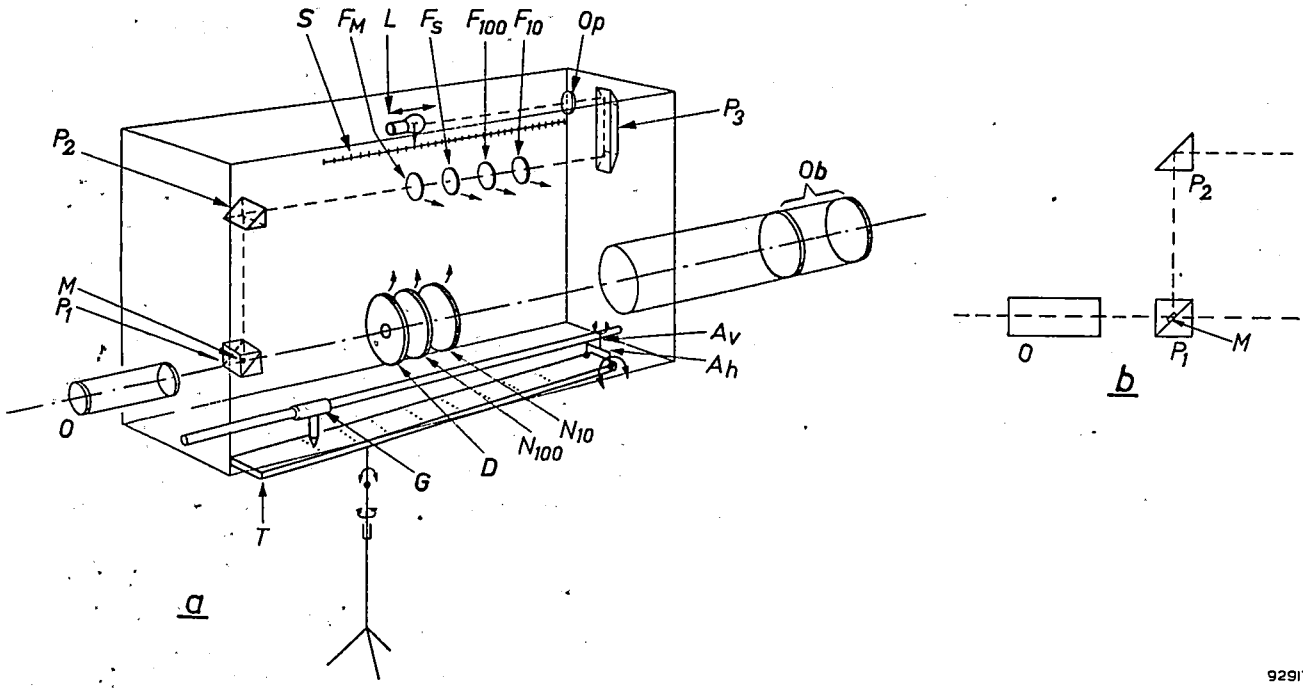


Fig. 4. *a*) Schematic representation of the visual luminance meter developed for road-lighting measurements. The objective *Ob* projects an image of the part of the road under investigation on to the mirror *M* (vapour-deposited on one of the prisms of P_1 — see detail in (*b*) — before these parts are cemented together). Through the eyepiece *O*, the image of the road can be sharply observed together with the mirror itself. The brightness of the comparison field from *M* can be adjusted until it is equal to that of the part of the road appearing immediately above the upper edge of *M*. This is done by varying the distance of the lamp *L* from the frosted-glass disc *Op* (to be read from a scale *S*). Via prisms P_2 and P_3 the brightness of *Op* is observed in *M*. The attenuation filters F_{10} and F_{100} can be interposed in the path of light in order to reduce the brightness observed in *M* by a factor of 10, 100 or 1000, thereby shifting the measuring range to lower luminances. Filters F_S and F_M serve for adapting the colour of the comparison field to a road lit by sodium or mercury-vapour lamps. With filters N_{10} and N_{100} the measuring range can be shifted to higher luminances. *D* is a diaphragm. The total measuring range is from 0.01 to 10^6 cd/m^2. A particular feature of the meter is its very small comparison field (the dimensions of *M* are 0.2×0.6 mm), which enables specific parts of the road to be measured even from a considerable distance. Another feature is the facility for accurately adjusting the optical axis of the meter with respect to the road. The meter can be turned about a horizontal axis A_h and a vertical axis A_v with respect to a hardened steel plate *T* fixed to the stand. By a certain procedure employing datum lamps placed on the road, a specific fixed position with respect to the road can be given. A sleeve *G*, which can be slid along a rod fixed to the meter, has a pin attached to it which rests on plate *T*. If *G* is pushed forward, the meter tilts slightly forward about A_h. The point of the road on which the meter is directed is entirely determined by the position of the pin of *G* on *T*. By calibration marks on the plate *T* it is possible to sight the meter on the point of the road where the luminance is to be measured. For this purpose *T* is illuminated, and an optical device with cross-wires is fitted to *G* for reading off the scale.

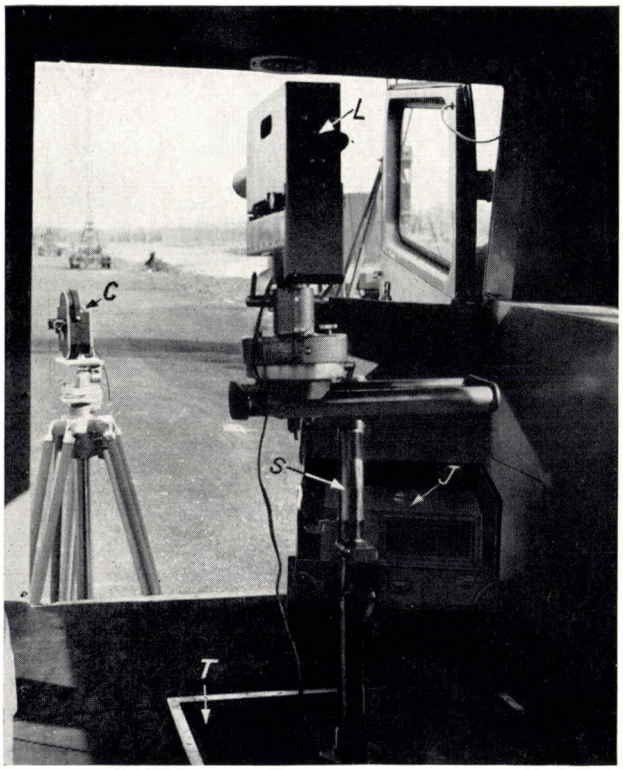
situation) are not, however, confined to the two types of surface mentioned, samples of several other road surfaces being mounted on low trolleys which can be moved to any position and hence included in the measurements. To make the tests less dependent upon the weather, water-sprinklers are situated at intervals of 20 m down the middle of the road.

The measurements and observations now being carried out in the open-air laboratory concern, apart from the luminance, the intensity of illumination on the road surfaces, the glare caused by direct light from the fittings and the visibility of objects on the road. The situations investigated, in addition to being measured photometrically, are recorded photographically as faithfully as possible.

Some of the instruments and apparatus used are contained in a specially equipped van, which also serves for carrying out measurements on existing road-lighting installations. The instruments employed for measuring the luminance and its distribution on the road surface are of two types: visual (subjective) instruments and instruments permitting objective measurements. One of the

most useful instruments is a luminance meter developed specially for road-surface measurements by the Eindhoven Lighting Laboratory. With this type of meter the luminance, seen from a fixed point where the meter is set up, is measured as a function of direction. Particularly on sections of the road surface far removed from the meter, a slight change in direction corresponds to a large displacement over the road surface. Consequently, one of the requirements imposed on the meter is that the direction of its optical axis should be capable of being adjusted and read off with great accuracy: for the azimuth setting the accuracy must be a few minutes of an arc, and even greater accuracy is required for adjusting the angle of inclination in the vertical plane. A schematic representation of the luminance meter is given in *fig. 4;* the optical principles of the meter and the measures adopted to ensure accurate adjustment are briefly described in the caption [1]). These measures are only effective, however, provided the meter is set up on firm ground. For this reason the meter on its stand can be let through a hatch in the floor of the van directly on to the road (*fig. 5*); this arrangement has the added advantage that instrument and observer are protected from wind and weather by the shelter of the van.

The data obtained from the tests performed in the open-air laboratory provide a welcome supplement to the knowledge gathered by experiment and experience over the years, and can play a great part in ensuring that the development of public lighting technique keeps pace with the rapid increase in the density of traffic on the roads.



92796

Fig. 5. Interior view (towards the rear) of the van (see also fig. 2) equipped with measuring instruments. The luminance meter L is mounted on a stand S, which is let down to the road through a hatch T in the floor of the van. J is a luxmeter, C a cine camera.

[1]) For a fuller description see J. B. de Boer, Lichttechnik **7**, 273-275 and 307-309, 1955, or Revue générale des Routes et des Aérodromes **25**, 83-90, 1955.

I. HAMMING and
J. F. T. van HEEMSKERCK VEECKENS.

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.

2470: M. J. Sparnaay: Zur Deutung der Hückel-Krafftschen Zusatzglieder in der Debye-Hückelschen Theorie der starken Elektrolyte (Z. phys. Chem. (Frankfurt a. M.) **10**, 156-160, 1957, No. 3-4). (On the explanation of the Hückel-Krafft extra terms in the Debye-Hückel theory of strong electrolytes; in German.)

In the classical theory of electrolyte solutions a quantity K^{-1} was introduced, having the dimension of length and usually interpreted as the mean distance between positive and negative ions constituting the solution. One of the assumptions of the theory was that the ions were considered as point charges. A modified theory which assumes that the ions are hydrated, and therefore occupy a certain volume, results in the quantity K^{-1} being replaced by K^{*-1} where, approximately,

$$K^{*-1} = K^{-1}\{1 + 2Bn(1-g)\}.$$

B is the volume of the hydrated ion, n is the number of ions per cm^3 of the solution and g is a factor containing the electric charges of the ions. The purpose of the present paper is to present a simple

theory for the understanding of g instead of the very complicated theory used hitherto. It is found that, generally, $g < 1$.

2471: M. J. Sparnaay: Non-equilibrium diffuse double-layer (Trans. Faraday Soc. **53**, 306-314, 1957, No. 3).

The diffuse part of the double layer in an electrolyte solution near an electrode is described by the Poisson-Boltzmann distribution. This distribution will undergo perturbation if an electric current is applied. Then the ions in the double layer will transport electricity. Two idealized cases of this non-equilibrium state of the diffuse part of the double layer are calculated. The case of an alternating current is also treated. It appears that deviations from the equilibrium distribution are nearly always small in practical cases.

2472: D. Hofman, J. A. Lely and J. Volger: The dielectric constant of SiC (Physica **23**, 236, 1957, No. 3).

Letter giving results of dielectric-constant measurements of very pure SiC crystals at 20 °K and 77 °K and at frequencies between 1 kc/s and 100 kc/s. It is concluded that $\varepsilon = 10.2 \pm 0.2$.

2473: P. C. van der Willigen and L. F. Defize: Das Lichtbogenschweissen von Stahl mit CO_2 als Schutzgas (Schweissen und Schneiden **9**, 50-59, 1957, No. 2). (Arc welding of steel with CO_2 as shielding gas; in German.)

See these abstracts No. **2394.**

2474: A. van Weel: Implications of phase precompensation in a television transmitter on the shape of the radiated signal (J. Brit. Instn. Rad. Engrs. **17**, 129-134, 1957, No. 2).

The smears after black-to-white transitions due to phase errors in the receiver can be compensated by an overshoot introduced by a phase-precompensating network in the video-frequency section of the transmitter. In a negative-modulation system this overshoot can only be accommodated in the available modulation space by raising the maximum-white level from 10 per cent to at least 20 per cent, which is equivalent to a loss of transmitted signal power of 30 per cent. Curvature of the modulation characteristic makes a complete compensation impossible. With a positive-modulation system, the output stage of the transmitter should be capable of delivering pulses, reaching to 22 per cent over and above the maximum-white level.

2475: R. Dijkstra and J. de Jonge: The ortho/para distribution of methylol groups in the reaction of phenol and formaldehyde II (Rec. Trav. chim. Pays-Bas **76**, 92-100, 1957, No. 1).

A potentiometric method is described for the quantitative determination of o-methylol phenol. The method is based on the acidic properties of the complex formed by o-methylol phenol and boric acid. In the reaction of phenol and formaldehyde, leading to the formation of mono-methylol compounds, 57-61% ortho substitution takes places. The distribution of the methylol groups over the ortho and para places of the phenolic nucleus was determined with the aid of the potentiometric method mentioned above.

2476: H. A. Klasens, A. H. Hoekstra and A. P. M. Cox: Ultraviolet fluorescence of some ternary silicates activated with lead (J. Electrochem. Soc. **104**, 93-100, 1957, No. 2).

The fluorescence and phase diagrams of the following Pb-activated ternary systems were investigated: $SrO\text{-}MgO\text{-}SiO_2$; $SrO\text{-}ZnO\text{-}SiO_2$; $BaO\text{-}MgO\text{-}SiO_2$; $BaO\text{-}ZnO\text{-}SiO_2$. Eleven new ternary silicates were identified, some of which proved to be very good base materials for Pb-activated phosphors with peak emissions ranging from 3000 to 4000 Å.

2477: H. Bremmer: Theoretische beschouwingen betreffende de atmosferische verstrooiing van radiogolven (T. Ned. Radiogenootschap **22**, 87-106, 1957, No. 2). (Theoretical considerations on the atmospheric scattering of radio waves; in Dutch.)

Theories of scatter propagation are dealt with. The importance of the Fourier spectrum connected with the spatial distribution of the local deviations of the dielectric constant is discussed. The influence of the assumed spatial auto-correlation function of these deviations is also put forward. Various consequences of the theory, such as fading and coherence properties, are briefly mentioned.

2478: J. A. Kok and M. M. G. Corbey: Testing the electric strength of liquid dielectric or insulating material (Appl. sci. Res. **B 6**, 285-295, 1957, No. 4).

Starting from the assumption that electric breakdown of liquid hydrocarbons takes place in general far below the intrinsic breakdown strength owing to irremovable and filterable contaminants migrating towards a place of maximum electric stress, where a conducting bridge may be formed, the authors attempt to combine the theory they propose with

earlier experimental results. Hydrocarbon oil of rather poor quality was purified by means of an electrostatic Cottrell-filter. This filter removed most of the foreign particles of high dielectric constant, by which the one-hour breakdown strength was increased from about 10 kV/mm to 60 kV/mm.

2479: J. H. Uhlenbroek, M. J. Koopmans and H. O. Huisman: Investigations on agricultural fungicides, I. Trichloromethyl thiolsulphonates (Rec. Trav. chim. Pays-Bas **76**, 129-146, 1957, No. 2).

The preparation and the biological properties of a number of strongly fungitoxic compounds of the general type RSO_2SCCl_3 (trichloromethyl thiolsulphonates) are described in this paper. The phytotoxic character of these compounds may be lowered considerably by suitable substitution in the group R.

2480: J. F. Carrière: Bijzondere elektronenbuizen voor sommige impulstechnieken (Ned. T. Natuurk. **23**, 97-115, 1957, No. 4). (Special electron tubes for use in certain pulse applications; in Dutch.)

Properties and limitations of various types of electron tubes for counting purposes. Among the tubes dealt with are gas-filled tubes, vacuum tubes, tubes making use of secondary emission, cathode-ray tubes (beam-switching tubes) and the trochotron (or magnetron-type) beam-switching tube.

2481: J. P. L. Bots: Resorption of cholecalciferol from an intramuscular oil depot (Rec. Trav. chim. Pays-Bas **76**, 209-212, 1957, No. 3).

By injecting chickens intramuscularly with radioactive cholecalciferol (vitamin D_3) in oil and determinating the radioactivity still present after various times, it is shown that an intramuscular depot of vitamin D_3 is only slowly absorbed in the course of at least seven months.

A 1: A. Rabenau: Perowskit- und Fluoritphasen in den Systemen ZrO_2-$LaO_{1.5}$-MgO und ZrO_2-$LaO_{1.5}$-CaO (Z. anorg. allgem. Chemie **288**, 221-234, 1956, No. 3-4). (Perovskite and fluorite phases in the systems ZrO_2-$LaO_{1.5}$-MgO and ZrO_2-$LaO_{1.5}$-CaO; in German.)

In the system ZrO_2-$LaO_{1.5}$-MgO a ternary compound has been found having the composition $La(Zr_{0.5}Mg_{0.5})O_3$. This compound has a perovskite structure and a narrow range of homogeneity. A compound with the same composition and structure was found in the CaO system: $La(Zr_{0.5}Ca_{0.5})O_3$. Here it forms a complete range of mixed crystals with $CaZrO_3$. The occurrence of cubic fluorite phases in the ternary systems has been investigated. The fluorite phases of ZrO_2 with the oxides MgO and CaO are not miscible with the ZrO_2-$LaO_{1.5}$ fluorite phase. On the other hand the latter with a composition of $ZrO_2 : LaO_{1.5} = 1 : 1$ ($La_2Zr_2O_7$) can take up to about 50 mol% of MgO. This behaviour is explained by the fact that with this composition a special structure occurs, as is shown by the occurrence of superstructure lines. A compound of the pyrochlore type is probably involved.

R 319: B. van der Veen: On the angle between wave front and displacement of plane acoustic waves in quartz (Philips Res. Rep. **12**, 273-280, 1957, No. 4).

Piezoelectric vibrating X-cut bars of quartz are generally made rectangular. Following a suggestion of C. Franx it is shown that in general a non-rectangular shape can give zero coupling to unwanted transversal vibrations. In this case it is necessary to compute the angle between the wave front and the displacement of plane acoustic waves in quartz.

R 320: J. Bloem and F. A. Kröger: Interstitial diffusion of copper in PbS single crystals (Philips Res. Rep. **12**, 281-302, 1957, No. 4).

At temperatures $100 < T < 500$ °C, Cu can diffuse rapidly into PbS via interstitial sites. Existing cation vacancies are filled by copper, while copper may also be bound by sulphur at internal surfaces (cracks, dislocations). The diffusion co-efficient follows the relation $D_{Cu} = 5.10^{-3} \exp(-7130/RT)$ $cm^2\ sec^{-1}$. Interstitial copper gives rise to donor levels lying $\sim$0.02 eV below the conduction band, and thus causes n-type conductivity. In the presence of sulphur in the atmosphere (H_2S) copper does not enter the crystal; interstitial copper already present in the crystal may be drawn out when the crystal is heated in such an atmosphere. The effect is attributed to a lowering of the thermodynamic potential of copper by the formation of Cu_2S or CuS.

R 321: J. Bloem and F. A. Kröger: Interstitial diffusion of nickel in PbS single crystals (Philips Res. Rep. **12**, 303-308, 1957, No. 4).

Under reducing conditions, nickel may penetrate into PbS crystals at temperatures $T < 500$ °C at which the self-diffusion in PbS is negligible; it diminishes p-type conductivity and may cause n-type conductivity with donors of a depth $E \sim 0.03$ eV. The diffusion probably takes place via the

interlattice with a diffusion constant $D_{Ni} = 17.8 \exp(-22000/RT)$ cm^{2}sec^{-1}. Under sulphurizing conditions (H_2S) the nickel can be drawn out of the crystal again and the original conductivity restored.

R 322: C. Z. van Doorn: Anisotropy of colour centres in alkali halides (Philips Res. Rep. **12**, 309-323, 1957, No. 4).

Additively coloured crystals of KCl and NaCl, which have been irradiated in the F-band at room temperature, show, in addition to the F-band, absorption bands which are due to M- and R-centres. Irradiation in these bands with polarized light at 77 °K produces an unpolarized infrared emission due to F-centres and two infrared emission bands, due to M- and R-centres, showing polarization. The degree and directional properties of this polarization depend on the type of centre and method of excitation. Irradiation with polarized light in the F-band also induces dichroism in the F- and M-bands. Models for the centres are discussed which may explain these phenomena.

R 323: M. T. Vlaardingerbroek, K. S. Knol and P. A. H. Hart: Measurements of noisy fourpoles at microwave frequencies (Philips Res. Rep. **12**, 324-332, 1957, No. 4).

A new method of measuring the characteristic noise quantities of noisy linear fourpoles is described. Applied to microwave frequencies this method is very simple when the fourpole is matched to the characteristic impedance of the waveguide, which can always be achieved. The method is applied to a microwave triode amplifier.

R 324: J. L. Meijering: Residual entropy of ice, and related combinatorial problems (Philips Res. Rep. **12**, 333-350, 1957, No. 4).

The theoretical residual entropy of disordered ice is about 5% higher than Pauling's value $R \ln (3/2)$. For the corresponding quadratic lattice the difference is 6%. Here a persistency effect is found, which in $Ag_2H_3IO_6$ must give rise to unequal a-priori probabilities of the dipole orientations. The entropy computed for the graphite-type honey-comb lattice is 20% higher than that found by Lipscomb. This difference is related to a boundary effect. It is shown that Chang's expression for the entropy of diatomic molecules in a lattice can be derived directly with Pauling's method.

R 325: J. H. N. van Vucht: Some data about the system cerium-thorium (Philips Res. Rep. **12**, 351-354, 1957, No. 4).

Metallographic and X-ray diffraction experiments show a complete mutual solubility of cerium and thorium in the face-centred cubic phase. The lattice parameters deviate markedly from Vegard's law in the negative sense.

R 326: K. F. Niessen: Influence of foreign ions on the critical field strength of an antiferromagnetic (Philips Res. Rep. **12**, 355-363, 1957, No. 4).

The critical field strength is calculated for an anti-ferromagnetic in each of whose sublattices a number of original magnetic ions has been replaced by foreign magnetic ions with different magnetic moments or simply by non-magnetic ions. Especially in the latter case the change in the critical field strength can be given by a simple formula.

R 327: L. J. van der Pauw: Analysis of the photoconductance in silicon (Philips Res. Rep. **12**, 364-376, 1957, No. 4).

A general formula is derived for the change in voltage drop across a semiconductor sample, caused by illumination. It is shown that this formula is a convenient starting point for analysing photoconductive experiments. In particular it is shown that in relevant cases white light may be used without difficulty. The analysis of the low-frequency phase shift between the voltage drop and the light signal as a function of the thickness of the sample yields the values of the bulk lifetime, the surface-recombination velocity and the drift mobility of the minority charge carriers. Also the properties of traps can be determined. Finally the results of some measurements at room temperature are discussed.

VOLUME 19, 1957/58, No. 7-8 pp. 209-244 Published 10th February 1958

Philips Technical Review

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

CRYSTAL-ORIENTED FERROXPLANA

by A. L. STUIJTS and H. P. J. WIJN.

621.318.13:538.213

The magnetically soft materials "ferroxplana", on which an article recently appeared in this Review, have a permeability which remains constant up to frequencies far above 100 Mc/s. It appears that the crystals of these materials can be aligned. As a result the permeability is appreciably increased, with only a slight drop in the limiting frequency. Moreover, this produces a material with an anisotropic permeability, for which there are special applications.

Introduction

A short time ago some new groups of ferromagnetic oxides were discovered in the Philips Laboratory, Eindhoven, which, like the earlier described material ferroxdure, possess a hexagonal crystalline structure. These materials were described in an article in this Review [1]), which we shall henceforth refer to as I. The composition of these compounds can be represented in a triangle diagram, the vertices of which are formed by three oxides. This triangle is shown in *fig. 1*. Me represents a divalent ion from the series Mn, Fe, Co, Ni, Cu, Zn, Mg, or a mixture of these ions. On the sides of the triangle are found the compound **B** ($BaFe_2O_4$), the group of compounds **S** ($MeFe_2O_4$, which includes ferroxcube) and the compound **M** ($BaFe_{12}O_{19}$, main constituent of ferroxdure). The new groups of compounds discussed in I are represented in the triangle by the points **W** ($BaMe_2Fe_{16}O_{27}$), **Y** ($Ba_2Me_2Fe_{12}O_{22}$) and **Z** ($Ba_3Me_2Fe_{24}O_{41}$). Later, the new groups of compounds **X** and **U** were discovered [2]). If, in the **Z** group, for example, the Co ion is chosen for Me, this compound can be briefly denoted by Co_2Z, indicating that the structural unit (this is not the unit cell, see I, ref. [7]) and p. 153) contains two Co ions.

As in the case of ferroxcube and ferroxdure, these materials are prepared by sintering at high tempera-

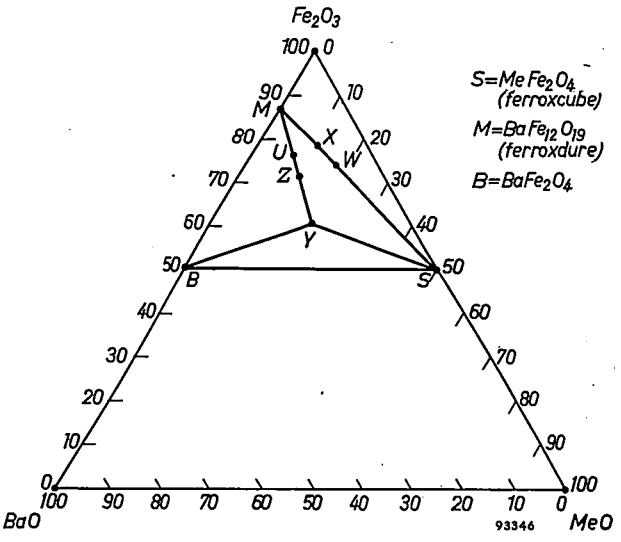


Fig. 1. The composition of the materials **W**, **Y**, **Z**, **X** and **U** can be represented by points in the triangle BaO-MeO-Fe_2O_3. **B**, **S** and **M** represent the known materials $BaFe_2O_4$, ferroxcube and ferroxdure, respectively.

[1]) G. H. Jonker, H. P. J. Wijn and P. B. Braun, Ferroxplana, hexagonal ferromagnetic iron-oxide compounds for very high frequencies, Philips tech. Rev. **18**, 145-154, 1956/57, (No. 6); referred to in this article as I.

[2]) See P. B. Braun, thesis, Amsterdam 1956, also Philips Res. Rep. **12**, 491-548, 1957 (No. 6); G. H. Jonker, Ferrimagnetic iron oxide compounds with hexagonal crystal structure, paper presented in June 1957 to 16e Congrès intern. Chimie pure et appl., Paris.

tures, which gives rise to a reaction between the constituent oxides and the atmosphere. The ceramic product so produced has a high resistivity (10^4-10^{10} ohm cm).

It has been found that in all compounds of the Y group, and also in those of the Z and W groups which contain more than a certain content of Co, the crystals exhibit a preferred plane for the magnetization; these compounds have been given the name ferroxplana. This preferred plane is perpendicular to the (hexagonal) *c*-axis, which we have called the "abhorred" direction. The magnetic anisotropy for rotations out of this plane is generally large, which is to say that the magnetization is very stiffly bound to this plane, and therefore the permeability in the direction perpendicular to the preferred plane is small. However, the magnetization can fairly easily be rotated in the preferred plane itself (relatively small anisotropy for rotations in the preferred plane) so that the permeability in directions parallel to this plane can be much higher [3]). In polycrystalline material, in which the preferred planes lie in arbitrary directions, there is then a certain average isotropic permeability. It was shown in I that, owing to the markedly differing anisotropies of the crystals, the product of the permeability and the frequency above which the losses show a sharp increase (limiting frequency) can attain a much higher value than with ferroxcube. For example, in the case of Co_2Z the permeability is found to be about 10 at a limiting frequency of about 400 Mc/s, while ferroxcube 4E with a permeability of 12 has a limiting frequency of about 90 Mc/s.

In this article we shall first discuss how the permeability of ferroxplana can be influenced by an anisotropic distribution of the orientations of the crystallites in the material (texture). We shall then describe a method of crystal alignment differing from a method described for ferroxdure in an earlier article in this Review [4]), henceforth referred to as II. It will be shown that by the alignment of ferroxplana crystals the permeability of the material can be considerably increased in certain directions; the explanation of this phenomenon will be discussed. The limiting frequency is only slightly reduced by the alignment.

[3]) Where permeability and susceptibility are referred to, the *relative initial* permeability and susceptibility are always meant (in μ_0 units in the rationalized Giorgi system), i.e. the value measured in the demagnetized state upon application of a very weak magnetic field.

[4]) A. L. Stuijts, G. W. Rathenau and G. H. Weber, Ferroxdure II and III, anisotropic permanent magnet materials, Philips tech. Rev. **16**, 141-147, 1954/55; referred to in this article as II.

Anisotropic ferroxplana materials

The most anisotropic material is the single crystal. The anisotropic magnetic properties of a ferroxplana single crystal are exemplified in *fig. 2*, where

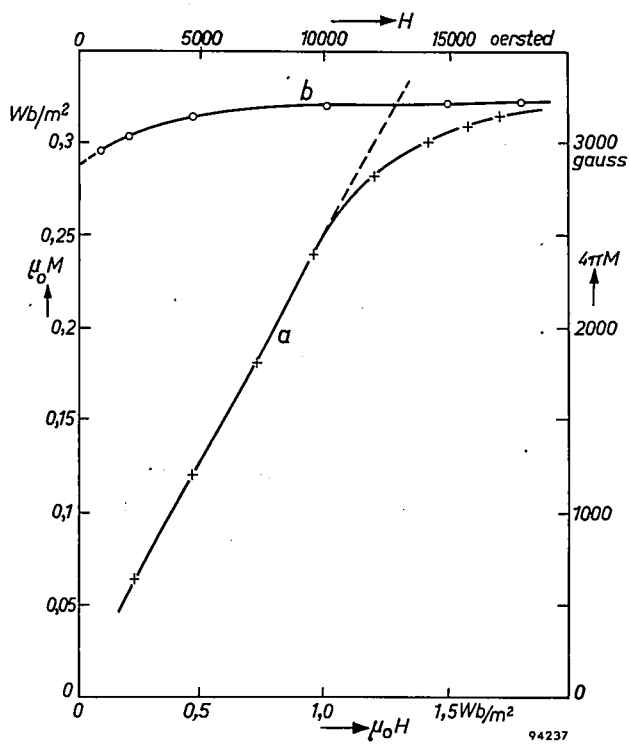


Fig. 2. The magnetization M of a Co_2Z single crystal as a function of the applied magnetic field H; *a*) preferred plane at right angles to H; *b*) preferred plane parallel to H.

the magnetization M for a single crystal of Co_2Z is plotted as a function of the applied induction $\mu_0 H$ in a direction perpendicular to the preferred plane (curve *a*) or parallel thereto (curve *b*). It can be seen from the figure that the induction $\mu_0 H$ needed to saturate the material is in the first case about 1.2 Wb/m² (field strength 12 000 oersteds), while in the second case a much smaller induction suffices for saturation.

In a normal polycrystalline specimen, the crystals have a random orientation, and therefore the specimen possesses isotropic magnetic properties. Textures are possible, however, in which anisotropic magnetic properties occur in the polycrystalline specimen. These textures are the following:

1) The basal planes of the crystals, which coincide with the preferred planes, are all parallel to one line, as illustrated in the upper half of *fig. 3a* (fan texture). In the lower half of this figure the orientation of the crystals is illustrated as they are seen when a cut is made perpendicular to the common line of the basal planes (the *c*-axes of all crystals

lie in the plane of this figure). It should be borne in mind in this connection that the crystals are generally in the form of platelets, with the smallest dimension in the direction of the *c*-axis. The permeability will now be greater in the direction of the common line than in the isotropic material, whereas the permeability in the perpendicular directions will be smaller.

2) All basal planes are mutually parallel, as illustrated in fig. 3*b* (foliate texture). In this case the permeability in all directions parallel to the basal planes is greater than that in the perpendicular direction.

As is known, the permeability is caused by the fact that an applied field $\vec{H}$ changes the direction of the spontaneous magnetization $\vec{M}_s$ in each Weiss domain if the couple $\mu_0\vec{H}\times\vec{M}_s$ differs from zero. In ferroxplana the magnetization is so strongly bound to the preferred plane, and hence the rotation of $\vec{M}_s$ out of this plane is so slight, that this rotation makes a negligible contribution to the permeability. For this reason, the only contribution to the permeability is made by the component of $\vec{H}$ that lies in the preferred plane of a given crystal. In a specimen in which the crystals are isotropically oriented, the permeability is therefore smaller than in a specimen with aligned crystals, in which $\vec{H}$ is made to lie parallel to the preferred planes.

For the case where there is no magnetic interaction between the crystals, the alignment causes the rotational susceptibility $\chi_i = \mu_i - 1$ to increase by a factor 1.5. The interaction, however, is by no means negligible, as appears from the idealized case illustrated in *fig. 4*. This figure shows a number of crystals of a specimen, with the preferred plane of the centre crystal assumed to be parallel to the hatching and perpendicular to the plane of the drawing. The preferred planes of all the other crystals are assumed to be parallel to the plane of the drawing. In this case the lines of force of the applied field must bend to the left and right around the "crosswise" crystal, so that this crystal has a strong demagnetizing influence. As a result the permeability is smaller than that of a specimen in which the preferred planes of all crystals are in parallel alignment.

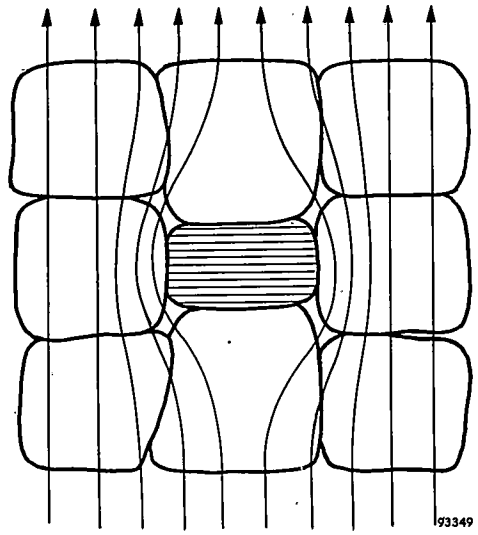


Fig. 4. Demagnetization due to a wrongly oriented crystal in ferroxplana. The preferred planes of the surrounding crystals lie in the plane of the drawing; that of the central crystal is perpendicular to the plane of the drawing and parallel to the hatching. The lines of force must bend round the "crosswise" crystal.

It will be shown below that both mechanisms here described are effective, and that therefore the permeability can be increased by the process of alignment by a factor greater than 1.5.

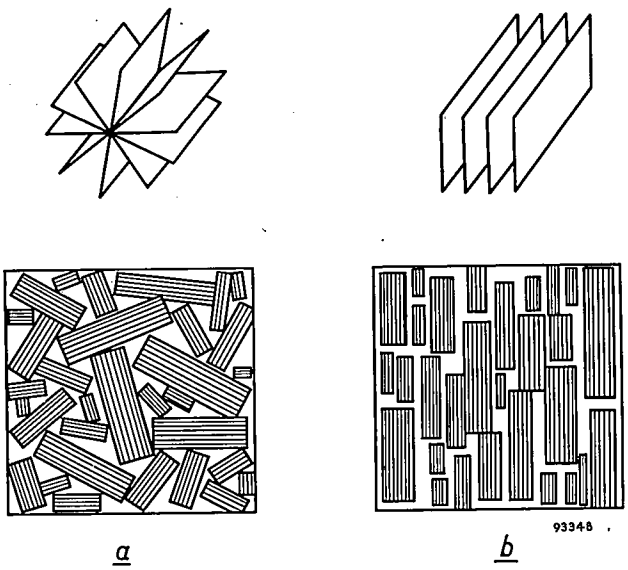


Fig. 3. Anisotropic ferroxplana. *Above*: basal planes in perspective; *below*: orientation of crystals in a section perpendicular to the basal planes. *a*) Fan texture; *b*) foliate texture.

Methods of aligning crystals of magnetic materials

There are various known methods of producing a magnetic material with a preferred orientation.

In the case of the metallic magnetically soft materials fernicube (50Ni-50Fe) and "grain-oriented" silicon-iron (3% Si, balance Fe) a crystalline texture is obtained by rolling and subsequent recrystallization. With fernicube the preferred directions of the crystals are aligned at right angles to the direction of rolling, so that in the latter direction

the relation between the permeability and the applied field becomes more linear; with silicon-iron the permeability is increased in the direction of rolling, because the preferred directions of the crystals are aligned parallel to this direction.

Rolling cannot be applied in the same way to ceramic materials because they cannot be plastically deformed. Use might be made, however, of the special form of the particles (which, as stated, are mostly flat with the smallest dimension in the c-direction). It has been found possible with ferroxdure to obtain a crystalline texture by compressing the powdered material in a steel tube and then rolling. The effect is slight, however, and of the same order as the texture obtained by pressing the preheated powder in a die.

With some magnetically hard materials (ferroxdure, MnBi) the fact that the magnetization is strongly bound to a preferred direction can be put to very good use for aligning the crystals and thereby substantially increasing the energy product $(BH)_{max}$. (For ferroxdure, see II.) In ferroxplana the magnetization is strongly bound not to a preferred direction but to a preferred plane (as a consequence of which ferroxplana is magnetically soft). Here, too, it has been found possible to align the crystals by making use of a magnetic field. To do this we start with a uniform magnetic field $\vec{H}$ in which a number of ferroxplana crystals are placed, each of which, it is assumed, are free to rotate. The magnetization $\vec{M}_s$ of each particle is drawn in the direction of $\vec{H}$ by a couple $\mu_0 \vec{H} \times \vec{M}_s = \mu_0 H M_s \sin \alpha$ (see *fig. 5*), where α is the angle between the magnetization $\vec{M}_s$ and the field direction. The crystal anisotropy field H^A binds the spontaneous magnetization $\vec{M}_s$ to the preferred plane with a couple which proves to be equal to

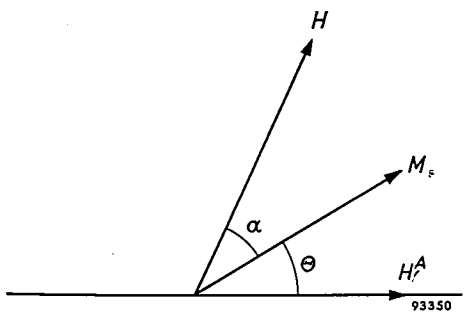


Fig. 5. Rotation of a ferroxplana crystal in an external field H. Owing to the couples $\mu_0 H M_s \sin \alpha$ and $\mu_0 H^A M_s \sin \Theta \cos \Theta$ acting on M_s, the preferred plane tends to take up a position parallel to H.

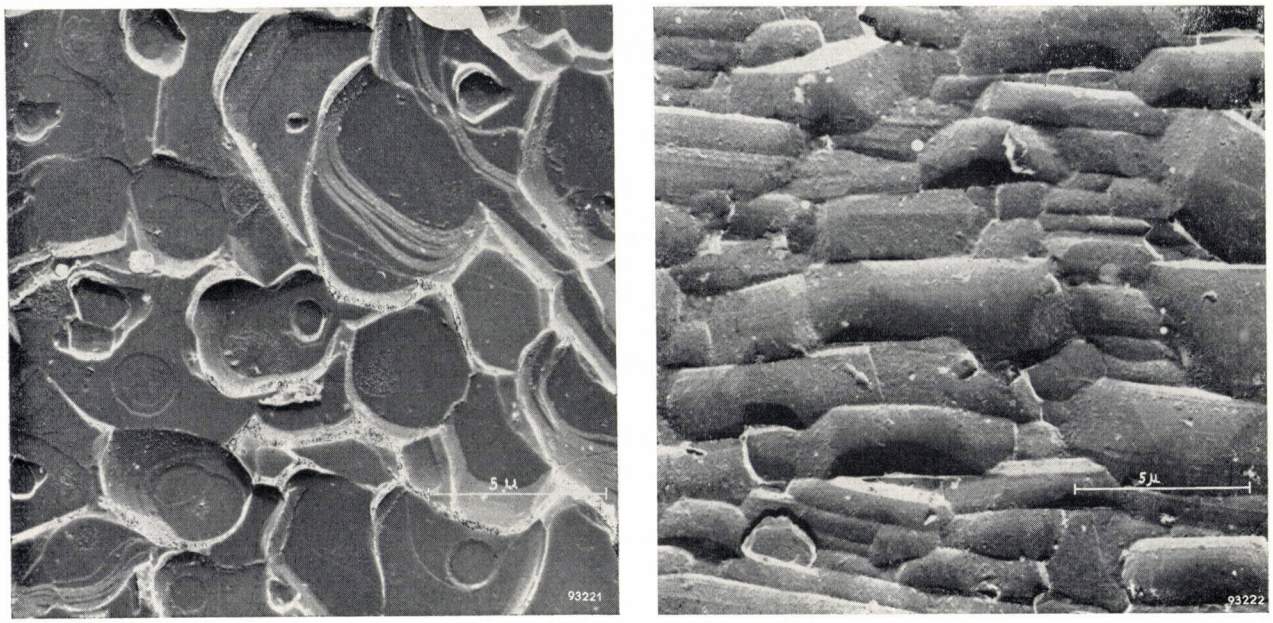
$\mu_0 \vec{H}^A \times \vec{M}_s = \mu_0 H^A M_s \sin \Theta \cos \Theta$, where Θ is the angle between the magnetization and the preferred plane. If the particle is free to move, it will then rotate until in a state of equilibrium $\alpha = \Theta = 0$, the preferred plane then being parallel to H. The plane itself can be in an arbitrary direction so that we have obtained here a fan texture.

To produce the foliate texture we subject the powder alternately to two magnetic fields whose lines of force are at right angles to each other, for example by energizing alternately two crossed electromagnets. The only stable state for the orientation of the particles is now that in which the preferred plane is parallel to both field directions, i.e. the foliate texture. Naturally, this state of equilibrium occurs only if the motion of the particles is damped, but in practice this condition is always satisfied. Instead of using two alternate, mutually perpendicular fields, we can use a magnetic field continuously varying in direction but lying in one plane, that is a rotating magnetic field which can, for example, be produced by mechanically rotating a yoke magnet. A rotating magnetic field can also be produced with a stationary magnet with the aid of the three phases of the electric mains. An electromagnet on this principle has been developed by U. Enz in the Philips Research Laboratories in Eindhoven, and has been found very suitable for alignment processes.

To produce good results a number of conditions must be fulfilled. The basic material is a magnetic powder that is obtained by pre-sintering and subsequent grinding. In the first place the preferred planes of the crystals in one particle must be mutually parallel (see II). Use is preferably made of a suspension of the powder in a liquid in order to produce a state in which the particles can readily be rotated. In order to be able to make a well finished ceramic product the particles, after orientation, should be stacked as compactly as possible. For this reason the process of aligning the particles is combined with a pressing treatment. This means, however, that the particles obstruct each other, so that an ideal state of alignment can never be achieved.

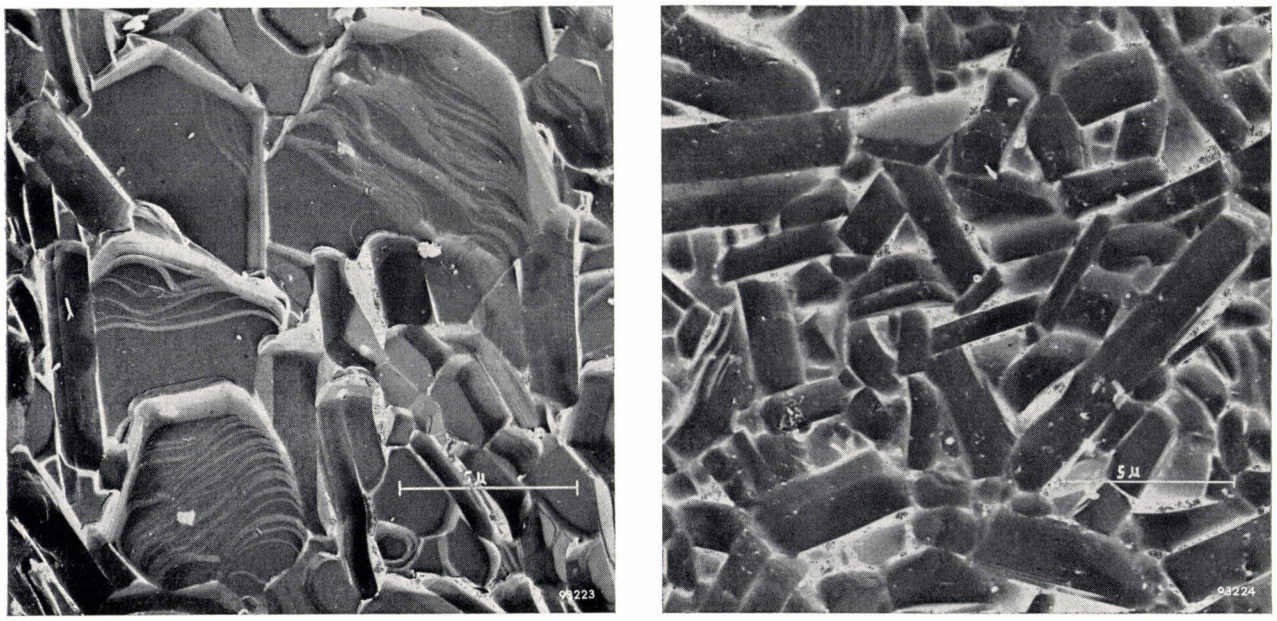
Moreover, the method can only be employed if the principal anisotropy field H^A is not too weak. For materials with a composition with which H^A is weak, as in certain mixed crystals of Co_2Z with other Me_2Z compounds, the result is therefore poor. It may be said of these materials that they show the effect typical of ferroxplana to a much lesser degree.

Fig. 6 shows two micrographs taken with an



a *b*

Fig. 6. Electron-microphotographs of aligned ferroxplana with foliate texture. In (*a*) the preferred planes are parallel to the plane of the paper, in (*b*) they are perpendicular thereto.



a *b*

Fig. 7. Photographs of a specimen with fan texture; the direction to which the preferred planes are parallel lies in (*a*) in the plane of the paper (roughly vertical) and in (*b*) perpendicular to the plane of the paper.

electron microscope; the foliate texture of the crystals is here clearly visible. In fig. 6*a* the preferred planes (the basal planes of the crystals) lie parallel to the plane of the paper; in fig. 6*b* they lie perpendicular thereto. The latter micrograph gives an idea of the platelet shape of the crystals. *Fig.* 7 shows two micrographs of a fan texture. In fig. 7*a* the direction to which all preferred planes are parallel lies in the plane of the paper; in fig. 7*b* it is perpendicular thereto.

During sintering, anisotropic shrinkage occurs in aligned ferroxplana as it does in aligned ferroxdure. In the direction of the hexagonal axis of the crystals, the shrinkage is much greater than in the direction perpendicular thereto, as can be seen in *fig.* 8.

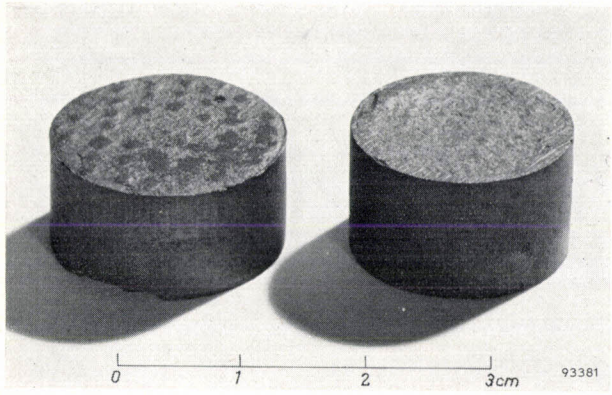


Fig. 8. Anisotropic shrinkage occurs during the sintering of an aligned pressed material. *Right:* an unaligned specimen; *left:* an aligned specimen with foliate texture. In the latter case there is considerable shrinkage in the direction of the *c*-axes of the crystals, but less at right angles to that direction. Before sintering, the dimensions of both specimens were identical.

Properties of aligned ferroxplana

The degree of alignment

The crystal-anisotropy energy in a single crystal can be found by determining the magnetization curve both in the preferred direction and in the "abhorred" direction, as in fig. 2. The anisotropy energy is equal to the area between the two curves.

For a material with a foliate texture it is possible, since the orientation of the preferred plane is known, to measure the magnetization curves in the "abhorred" direction as well as parallel to the preferred plane. If we neglect the anisotropy in the plane, we can in this way also determine the crystal-anisotropy energy for an ideally aligned material. If the material is not ideally aligned, the two curves will be closer to each other and will bound an area which is a fraction f of the area between the curves for the single crystal. We shall call this fraction f the degree of alignment of the anisotropic material. It is not possible to determine this for materials with fan texture since the preferred planes are not parallel.

The degree of alignment of specimens so treated is found to reach 90% or more.

Permeability

To ascertain how the permeability depends upon the degree of alignment, we made a number of Co_2Z specimens with foliate texture. The degree of alignment can be varied from specimen to specimen by regulating the current producing the rotating field. The density was approximately the same for all specimens. In *fig. 9* the permeability of these specimens is plotted against the corresponding degree of alignment. The figure shows clearly the considerable gain in permeability, particularly with a high degree of alignment. A compound $Co_{0.8}Zn_{1.2}Z$, which has a permeability of 22 in the unaligned state, is found to have a permeability of 55 after alignment. There are indications that, with a more complete alignment than has hitherto been possible, an even higher permeability could be obtained.

During the measurement of the permeability the applied field is naturally taken parallel to the preferred planes. If we measure at right angles to this direction, we can expect a lower permeability than in the isotropic material. In fact, a permeability of only 2.5 was measured in this way on the Co_2Z specimen with $f = 0.91$. A value of 1.3 is calculated for an ideally aligned specimen.

Apart from its greater permeability, the new material also offers advantages, owing to the anisotropic character of the permeability, when used in a magnetic circuit with an air gap. In that case the component of the stray field perpendicular to the direction of the flux will be small owing to the low permeability in this direction. A useful application that comes to mind is, for example, around the air gap of a recording head in magnetic recorders, where the stray field should be concentrated in the smallest possible space; another useful application would be for obtaining a more uniform field in an air gap.

When aligned specimens are made with a fan texture a greater permeability is found in the most favourable direction than in a corresponding unaligned specimen. The increase, however, is always smaller than in the case of a foliate texture. We assume that this is due to the fact that the degree of alignment of a fan texture will always be smaller than that of a foliate texture. In the latter we find,

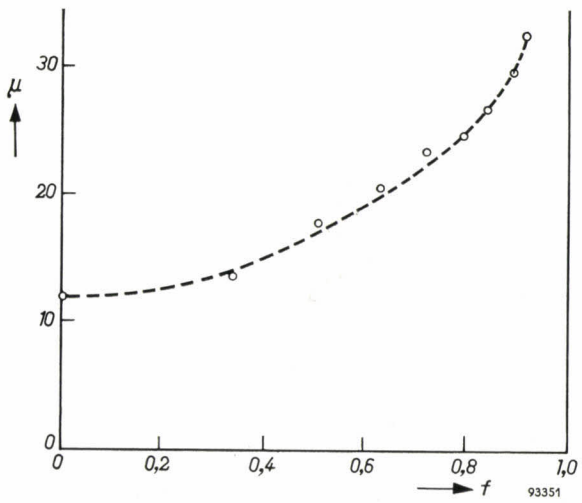


Fig. 9. The initial permeability μ of Co_2Z materials plotted against the degree of alignment f.

as in ferroxdure (see II) a marked increase in the alignment percentage during sintering at high temperature. This appears clearly from *Table I*; the degree of sintering is varied from specimen to specimen by changing the sintering temperature.

Table I. The influence of the sintering temperature on the degree of alignment f and the permeability μ of Co_2Z specimens.

Sintering temperature	f	μ
1180 °C	0.65	10.5
1200 °C	0.65	11.3
1220 °C	0.72	12.4
1240 °C	0.77	18.5
1260 °C	0.89	32.0

What probably happens is that the properly oriented crystals grow at the expense of the wrongly oriented ones, as described in II. It is not probable that the textural improvement needed for obtaining a high degree of alignment will occur in a material in which the basal planes are still in a fan-type distribution. As described, it is not possible to check this directly since f cannot be measured in a material with fan texture.

The stress anisotropy, too, can cause a difference in permeability between the materials with the two textures. Owing to the anisotropic coefficient of expansion of the crystals, more stress will arise during cooling in a material with fan texture than in one with foliate texture. If the magnetostriction is large, this is attended by a smaller permeability.

We have seen that, owing to the alignment, the permeability increases by much more than the factor of 1.5 expected for ideal alignment, owing to the fact that the external field can be applied parallel to the preferred planes. As stated, this additional gain is very probably attributable to the much smaller average internal demagnetization in aligned specimens; see fig. 4. We shall now examine this in more detail.

The internal demagnetization factor

Ferroxplana, like ferroxcube, is a sintered product containing pores. Consequently internal demagnetizing fields occur which reduce the permeability. A good impression of the demagnetization due to porosity can be obtained by measuring the so-called ideal curve. For this purpose an alternating magnetic field, whose amplitude gradually decreases from a high value to zero, is superimposed on a constant external magnetic field H. The curve which indicates the relation between the magnetization produced in this way and the intensity of the constant magnetic field is known as the ideal magnetization curve. If no demagnetization is present, the first part of the curve coincides with the magnetization axis, i.e. $(dH/dM)_{H=0} = 0$. Owing to demagnetization, the internal magnetic field H_i is not equal to the externally applied magnetic field H, and the ideal curve will make an angle a with the magnetization axis, determined by $\tan a = dH/dM = N_i$, where N_i is a so-called internal demagnetization factor (see *fig. 10*). It appears thaf between N_i and the porosity p of some types to

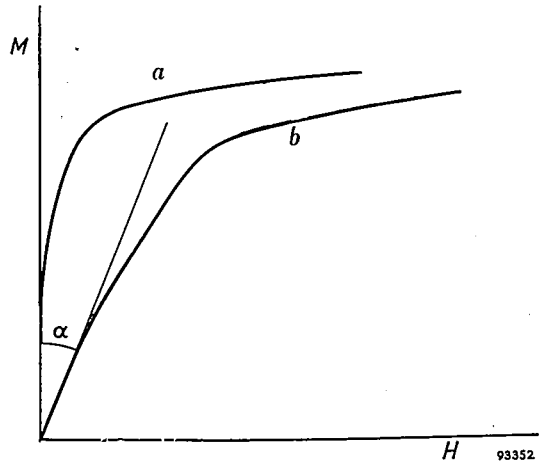


Fig. 10. The ideal magnetization curve. *a*) No internal demagnetization; *b*) internal demagnetization factor $N_i = \tan a$.

ferroxcube there exists a relation as represented by the curve *a* in *fig. 11*. As long as N_i is small, μ will not be limited by the internal demagnetization but by other anisotropies; in this region, therefore, small and large permeabilities are found side by side. When N_i is large, however, the internal demagnetization prevents the permeability from reaching a high value; indeed, the values found for μ decrease monotonically with increasing N_i.

If we determine in the same way the relation between N_i and p for various unaligned ferroxplana materials we get an entirely different picture, as can be seen from curve *b* in fig. 11. Even with very dense materials a fairly large N_i factor is found which, moreover, increases rapidly with the porosity; this curve as compared with that of ferroxcube is seen to have shifted some way to the left. The corresponding permeabilities are small. This anomalous behaviour can be understood by referring to fig. 4; the crystals lying cross-wise influence the lines of force in the same way as an air pore. The demagnetizing influence is greater still: in the surrounding crystals the lines of force must continue to run parallel to the preferred plane, i.e. in the plane of the

drawing. Thus, they can only bend round to the left and right and not forwards and backwards. It is understandable, then, that the internal demagnetization in a fairly dense, unaligned ferroxplana material is just as great as in a ferroxcube material having a porosity of about 30%.

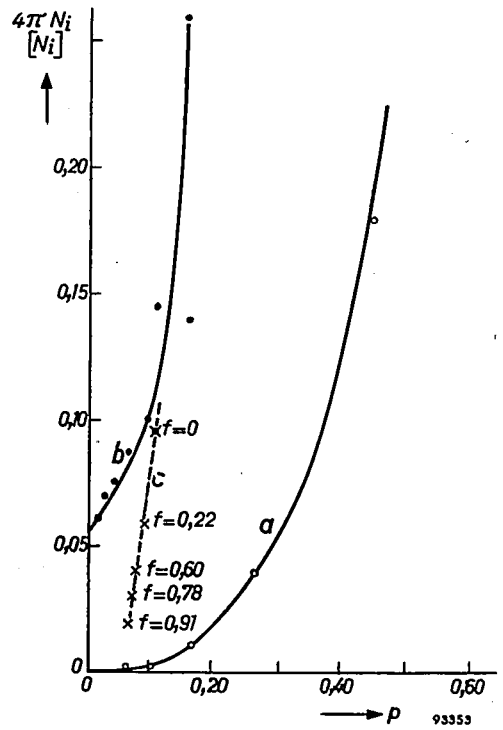


Fig. 11. The internal demagnetization factor N_i as a function of porosity p: *a*) for ferroxcube; *b*) for unaligned ferroxplana; *c*) for aligned Co_2Z with various degrees of alignment f.

We then measure N_i and p on aligned ferroxplana specimens, for which the relation between μ and f is given in fig. 9. The results of this measurement are shown in fig. 11, together with the appropriate degrees of alignment. We see that N_i decreases sharply with increasing degree of alignment f, just as expected: the ideal aligned state is approached at which, as in the case of ferroxcube materials, N_i is determined solely by the porosity. With well-aligned materials, the greatest gain in permeability is obtained by avoiding the demagnetizing influence of wrongly oriented crystals.

Frequency-dependence of the permeability

For application at very high frequencies a higher permeability is only of value if it does not entail a severe drop in the limiting frequency. In the case of specimens with a composition Co_2Z or $Co_{0.8}Zn_{1.2}Z$, therefore, the complex permeability $\mu = \mu' - j\mu''$ (μ'' is a measure of the losses: $\tan \delta = \mu''/\mu'$) was measured as a function of the frequency. The results are set out in *figs 12* and *13*; the curves for unaligned materials are shown for comparison. It is noticeable in both figures that the frequency

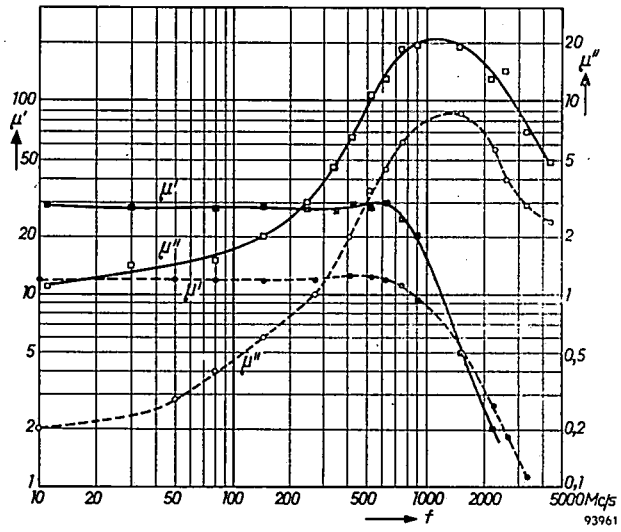


Fig. 12. The quantities μ' and μ'' as a function of frequency f; full curve for aligned Co_2Z, broken curve for unaligned Co_2Z.

at the peak of the μ'' curve is only slightly reduced by the alignment, in any case by a smaller factor than that by which the permeability is increased. The same applies to the limiting frequency f_r, if that is defined, for example, as the frequency at which $\tan \delta = 0.1$, i.e. the frequency of the point of intersection of the appertaining μ' and μ'' curves.

This can be explained as follows. Compared with an unaligned specimen an ideally aligned specimen, if the anisotropy field H_φ^A for rotations in the pre-

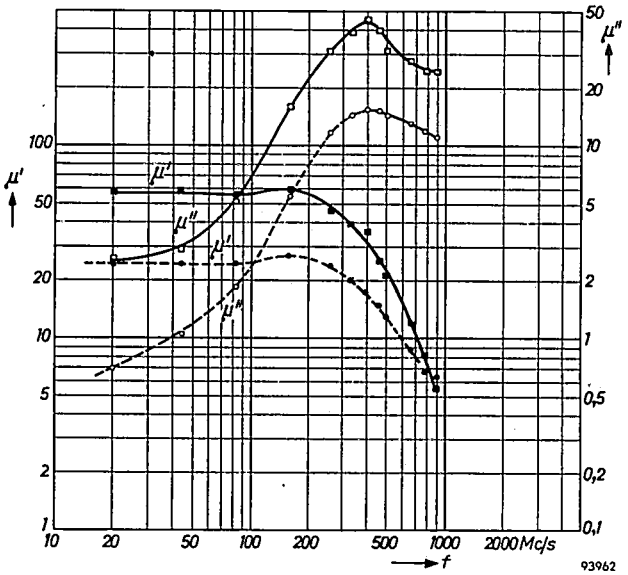


Fig. 13. The quantities μ' and μ'' as a function of frequency f for $Co_{0.8}Zn_{1.2}Z$; full curve for aligned material, broken curve for unaligned material.

ferred plane is unaltered, will show a susceptibility 1.5 times as large for a field applied parallel to the preferred plane as for a field perpendicular to this plane. As we have seen above, however, this anisotropy field H_φ^A becomes smaller by reducing the internal demagnetization; this additionally enhances the susceptibility. The large anisotropy field H_Θ^A for rotations *out of* the plane is not changed by the alignment, leaving out of account relatively small effects that might be caused by changes in the shape anisotropy. If we give all quantities of the unaligned material the suffix 1 and those of the aligned material the suffix 2, we can write:

$$\frac{\chi_2}{1.5\chi_1} = \frac{(H_\varphi^A)_1}{(H_\varphi^A)_2}.$$

From I we know that the limiting frequency f_r is proportional to $\sqrt{H_\Theta^A H_\varphi^A}$. Since $(H_\Theta^A)_1 = (H_\Theta^A)_2$, it follows that:

$$\frac{f_{r2}}{f_{r1}} = \sqrt{\frac{(H_\varphi^A)_2}{(H_\varphi^A)_1}} = \sqrt{\frac{1.5\chi_1}{\chi_2}} = \sqrt{\frac{1.5(\mu_1 - 1)}{\mu_2 - 1}}.$$

From the permeabilities μ_1 and μ_2 we calculate for Co_2Z the ratio $f_{r2}/f_{r1} = 0.77$; for $Co_{0.8}Zn_{1.2}Z$ we find $f_{r2}/f_{r1} = 0.80$. It can be seen from figs 12 and 13 that this is in good agreement with the experimental data.

Thus, by the alignment of the crystals of ferroxplana we can now obtain materials with a permeability of 30 to 50 and showing low losses up to frequencies of 200 and 100 Mc/s respectively.

Summary. Since the magnetization of ferroxplana materials is strongly bound to the preferred plane, the particles of a powdered specimen can be aligned in an external magnetic field. In a uniform field all preferred planes are parallel to the direction of the field (fan texture); in a rotating field all preferred planes are more or less mutually parallel (foliate texture). As a result the permeability becomes anisotropic and is large in certain directions; this gain in permeability is largely due to the avoidance, by alignment, of the strong demagnetizing effect of wrongly oriented crystals. In this way the permeability of some aligned specimens is increased by a factor of 2.5 to 3. For use at high frequencies it is important that the limiting frequency should not be greatly reduced; in aligned specimens this is about 0.8 times that in unaligned specimens. This effect can be explained quantitatively. The anisotropic nature of the permeability can also be put to good use in certain cases.

ANALYSIS OF THE GASEOUS CONTENTS OF SEALED CATHODE-RAY TUBES WITH THE AID OF THE OMEGATRON

544.4:621.385.832:537.534.3

Notwithstanding long and careful pumping and the use of highly active getters in cathode-ray tube manufacture, it is not feasible to evacuate below about 10^{-6} mm Hg. In addition to what is left in the tube, a certain amount of gas is liberated after the tube is sealed, both spontaneously from the various tube components and as a result of electron bombardment, from the fluorescent screen and any metal diaphragms present. The residual gas, i.e. the small fraction of the gas that is not bound by the getter, is undesirable in that it can shorten the life of the tube, *inter alia*, by poisoning the cathode [1]). If precautions are to be taken to reduce the amounts of harmful residual gases, it is necessary to know what gases are present in the tube and from which components of the tube they originate.

A useful aid to residual gas analysis has been found in the type of mass spectograph which Hipple and co-workers have called the omegatron [2]). The omegatron and cyclotron — widely different as they are in size — are analogous in that the charged particles are in both cases accelerated along (roughly) spiral paths. In the cyclotron the particles are accelerated twice per revolution in the gap between the two Dees [3]), whereas in the omegatron there is continuous acceleration in a homogeneous alternating electric field.

A schematic diagram of the omegatron is shown in *fig. 1*. A narrow beam of electrons passes from the cathode K to the electron collector A, parallel to a magnetic field of induction B. Above and below the beam, respectively, are the plates C_1 and C_2 which are connected to a radio-frequency generator G. The flat parallel rings D between these plates are so connected to tappings of a voltage divider P that there is an almost homogeneous alternating electrical field $E_0 \sin \omega t$ in the discharge space. The omegatron bulb is sealed via a glass tube to the vessel or electronic tube whose residual gas content is to be investigated.

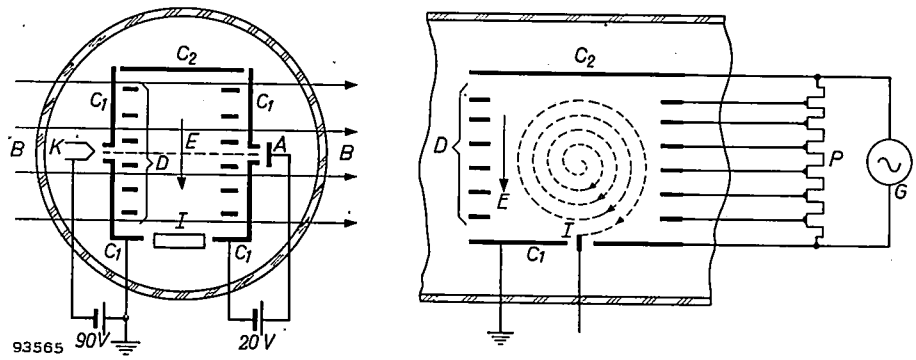


Fig. 1. Schematic diagram of an omegatron. K cathode (directly heated tungsten filament). A electron collector. C_1, C_2 plates to which the radio-frequency voltage from the generator G is applied. Flat metal rings D, connected to tappings of the voltage divider P, ensure homogeneity of the alternating electric field E in the discharge space. B constant magnetic induction. I ion collector.

The electron beam in the omegatron encounters the residual gas molecules, producing ions — almost entirely positive ions. The ions are accelerated by the R.F. field in a plane perpendicular to the magnetic field. If the angular frequency ω of the electric field satisfies the condition

$$\omega = \frac{e}{M} B \quad \ldots\ldots\ldots \quad (1)$$

(e = charge on the ion, M = mass of the ion), the resulting state of affairs is called *resonance*. This being the case, the ions describe a spiral of uniformly increasing radius (Archimedes spiral) and after making a certain number of revolutions are caught on a suitably positioned ion collector I. Ions having some other e/M ratio describe a different path and do not therefore reach the ion collector. The presence of the various ions can be successively inves-

[1]) Cf. Philips tech. Rev. **18**, 184, 1956/57 (No. 7).
[2]) H. Sommer, H. A. Thomas and J. A. Hipple, The measurement of e/M by cyclotron resonance, Phys. Rev. **82**, 697-702, 1951.
[3]) Cf. Philips tech. Rev. **12**, 67, 1950/51.

tigated by gradually varying the frequency of the electric field.

The omegatron has been used at Philips since 1953 for qualitative analysis of the residual gases in picture tubes [4]). With the type of omegatron used, it is possible with $B = 0.5$ Wb/m^2 to detect the presence of singly charged ions having a mass up to 50 (the most important residual ions have a mass in the region of 16; the heaviest are ions derived from CO_2 (mass 44) and certain hydrocarbons (masses from 44 to 58). Being of a convenient size (*fig. 2*), the omegatron can be connected by means of a short glass tube to the bulb of the picture tube before assembly; the omegatron remains connected with the picture tube while it undergoes assembly, processing and final sealing (*fig. 3*).

The other equipment is as follows. The radio-frequency voltage is drawn from a normal signal generator. The knob on the latter for adjustment of the frequency is slowly turned by a small motor so that the frequency varies through a certain range. A potentiometer, turned by the same shaft, supplies a uniformly increasing voltage; the latter is a measure of the frequency and produces a horizontal

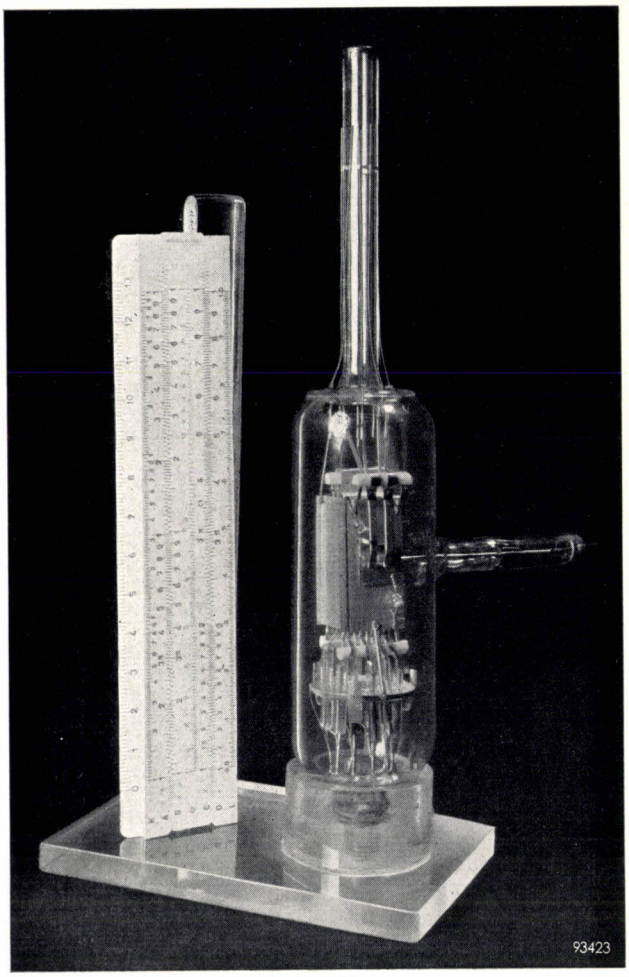


Fig. 2. One of the omegatrons employed.

[4]) Before this article went to press a paper on the same subject appeared elsewhere, namely: J. S. Wagener and P. T. Marth, Analysis of gases at very low pressures by using the omegatron spectrometer, J. appl. Phys. **28**, 1027-1030, 1957 (No. 9). The results reported agree broadly with those presented here.

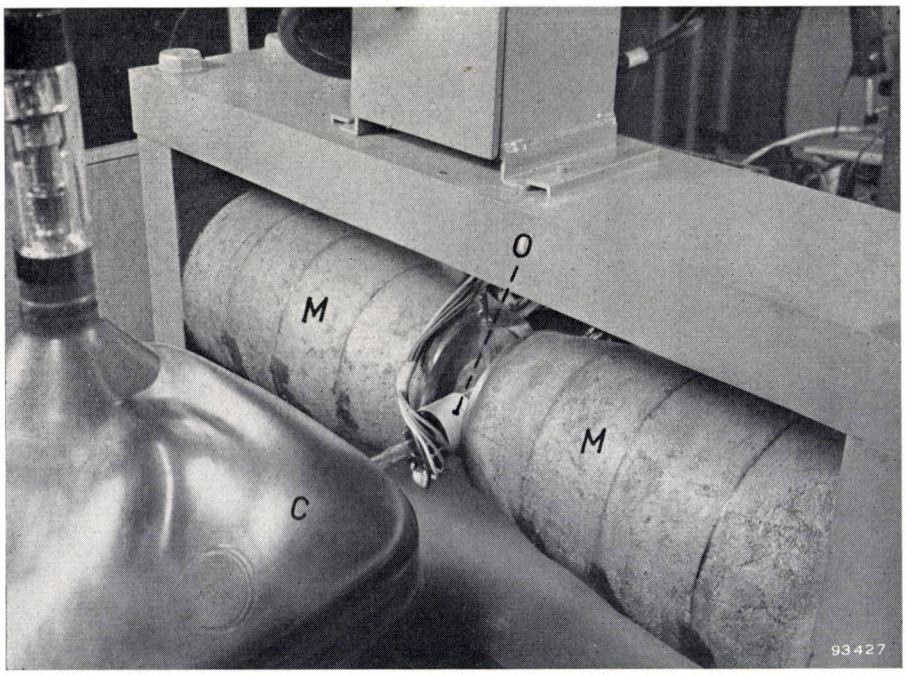


Fig. 3. An omegatron *O* connected to a picture tube *C* between the poles of a permanent magnet *M* which provides the constant magnetic field.

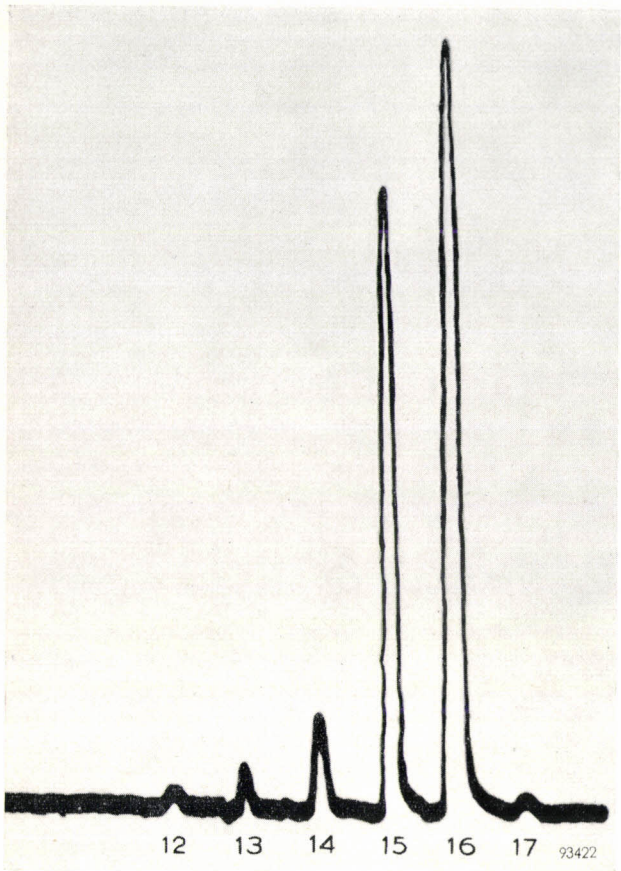


Fig. 4. A typical spectrogram. The peaks occur at frequencies which correspond, according to (1), to $M = 12, 13, 14, 15, 16$ and 17 (the ions carrying a single elementary charge e); these peaks are almost entirely attributable to C^+, CH^+, CH_2^+, CH_3^+, CH_4^+ and $C^{13}H_4^+$, respectively.

deflection on a cathode-ray oscilloscope. The current from the ion collector passes through a 10^{11} ohm resistor and the voltage across this resistor is amplified to give the vertical deflection. The light spot on the oscilloscope describes a curve which exhibits a peak at each frequency, the position of which, according to (1), corresponds to the mass M of the ions present. At present the heights of the pulses do not give an accurate measure of their relative concentration.

Some results of the investigation are given below. The principal residual gases that have been found in picture tubes are:

1) Methane. This gas was encountered in all the tubes examined. The high peak at $M = 16$ in the spectrogram (*fig. 4*) stems from CH_4^+ ions; the lower peaks at $M = 15, 14, 13, 12$ correspond to CH_4 fragments, viz. the ions CH_3^+, CH_2^+, CH^+ and C^+, respectively. The peak at $M = 17$ can be attributed to CH_4^+ with the carbon isotope C^{13}. The methane pressure falls with the passage of electrons in the picture tube, but rises again somewhat during periods of rest.
2) Argon. This gas was also found in all the tubes examined, but only after vaporization of the getter.
3) The various other residual gases met with depend upon the way in which the tube is manufactured and the materials used. The most common are hydrogen, carbon monoxide, carbon dioxide and hydrocarbons other than CH_4.

An example of a case in which CO and CO_2 can be expected are tubes in which the fluorescent screen is covered with a thin aluminium reflector ("metal backing") serving to reflect inwardly-radiated fluorescent light from the screen. If, deliberately, too little barium getter is vaporized in such a tube, there will be a rise in pressure subsequently when the tube is in use; this is brought about by CO and (to a lesser extent) by CO_2. These gases are in this case probably mainly produced when electrons collide with minute residual traces of the nitrocellulose film (the latter forms the underlayer for the aluminium and is removed by baking when it has served its purpose [5]).

Work is in progress with a view to making the omegatron and its associated apparatus suitable for quantitative analysis.

J. PEPER.

[5]) J. de Gier, Philips tech. Rev. **10**, 102, 1948/49.

A SIMPLE APPARATUS FOR CONTACT MICRORADIOGRAPHY BETWEEN 1.5 AND 5 kV

by B. COMBÉE and A. RECOURT.

778.33:621.386.1

Microradiography, the X-ray photography of microscopic specimens, can provide important information, unobtainable with an optical microscope. In principle, it is a fairly old technique, but its intricacy has hitherto debarred it from extensive use in hospitals and biological laboratories. With the appearance of the apparatus described in this article, the CMR 5, which is equipped with a small, sealed-off X-ray tube having a beryllium window only 50 microns thick, microradiography has now come within the range of every biologist or medical worker, and may even be employed for routine pathological examinations. The article below, after an introductory discussion of microradiography in general, describes the construction and uses of this new instrument. It appears that the resolution, which is now determined by the film, is 0.5 to 1 micron and hence not far behind what can be achieved visually with an optical microscope (about 0.3 micron).

Introduction

Among the methods of microscopic investigation, microradiography is rapidly gaining in significance and now occupies an important position along with optical and electron microscopy. In microradiography a radiograph is made by the "normal" method of shadow projection, and the X-ray image so obtained is examined with a magnifying glass or under a microscope. Thus it is not, as might at first be thought, microscopy with X-rays [1]), and therefore as regards resolution, the method does not offer the advantages that might be obtained with the very short wavelengths of an actual X-ray microscope [2]).

The fact that microradiography is nevertheless of great significance is due in the first place to the ability of X-rays to penetrate objects which are not transparent to light and electrons. Moreover, with our knowledge of the absorption of X-rays by matter, we can learn from a radiograph not merely the geometry of the object, but also something of its chemical composition. For example, it has proved possible in this way to determine the weight of histological and cytological [3]) structures and to carry out quantitative chemical analyses. Finally, by analogy with the staining of specimens for examination under an optical microscope, good use can be made of X-ray contrast media, which again offer quite different possibilities.

Before dealing with the apparatus which is the subject of this article, we shall examine briefly the laws governing the absorption of X-rays and the methods of projecting an X-ray shadow image on a film.

The absorption of X-rays; contrast

The attenuation which a parallel beam of monochromatic X-rays undergoes in matter is given by the formula:

$$I = I_0 \, e^{-\mu d}, \quad \ldots\ldots\ldots \quad (1)$$

in which I_0 is the initial intensity, I the intensity when the beam has traversed a layer of material of thickness d, and μ is the absorption coefficient. The way in which μ varies with the wavelength λ of the radiation and with the atomic number Z of the elements of which the absorbent material is composed is given for a wide range of wavelengths by the formula:

$$\mu = C\lambda^3 Z^4, \quad \ldots\ldots\ldots \quad (2)$$

in which C is a constant.

The dependence on λ is demonstrated in *fig. 1a*, which shows by how much an X-ray beam is attenu-

[1]) Although advances have been made in this field in recent years, the experimental X-ray microscopes so far developed are no better than optical microscopes as far as their resolving power is concerned. As regards magnification and field of view, they are greatly inferior. See P. Kirkpatrick and A. V. Baez, J. Opt. Soc. Amer. **38**, 766-774, 1948 and P. Kirkpatrick and H. H. Pattee, Advances biol. med. Phys. **3**, 247-283, 1953.
See also for X-ray optics: Y. Cauchois, Rev. Opt. **29**, 151-163, 1950 and M. Montel, Rev. Opt. **32**, 585-600, 1953; Optica Acta **1**, 117-126, 1954; C. R. Acad. Sci. Paris **239**, 39-41, 1954.

[2]) For a discussion of the dependence of the resolving power of microscopes on the wavelength of the light used, see, e.g., J. B. le Poole, Philips tech. Rev. **9**, 33-45, 1947/48.

[3]) *Histology* is the study of the structure of tissues, *cytology* the study of the structure of cells. Cells are the units of which tissues are composed.

ated in a layer of animal tissue 10 microns thick, in a 1 cm layer of air and in a layer of beryllium 50 microns thick; at a wavelength of 2 Å these layers are almost entirely transparent, whereas at 15 Å they absorb nearly all the radiation. Curves *2* and *3* of this figure reveal something of the very pronounced influence of Z. The atomic number of beryllium is 4, while the effective atomic number of air lies between 7 (nitrogen) and 8 (oxygen). If the layer of air were to be compressed to the density of beryllium, it would be only 7 microns thick, and yet it absorbs almost as strongly as the seven times thicker layer of beryllium.

The curve that represents absorption (or transmission) as a function of wavelength does not always have the continuous form shown in fig. 1*a*. In fig. 1*b* we see the absorption in a layer of photographic emulsion 5 microns thick plotted against the same range of wavelengths. At 3.7 and 7.0 Å there are abrupt discontinuities in the curve; these absorption edges, as they are called, occur at wavelengths which are characteristic of every element, and are caused here by silver and bromine.

At the wavelengths with which we are concerned, the attenuation of the X-ray beam is chiefly caused by the photoelectric effect. Where λ is large, electrons are liberated only from the outermost shell or shells of the atoms of the absorbent substance. As λ decreases, however, (i.e. as the energy of the X-ray quanta increases) a limit is reached where electrons can be liberated from the next, deeper shell, and the absorption suddenly becomes greater. These sharp discontinuities in the absorption curve are the above-mentioned absorption edges. They are denoted by the letter of the electron shell concerned, i.e. as K, L, M etc. (starting from the innermost shell of the atom). The curve in fig. 1*b* shows the L absorption edges of bromine and silver. The absorption edges of the elements making up the substances represented in fig. 1a lie at wavelengths longer than 15 Å.

Fig. 1*b* again illustrates the pronounced increase of absorption with atomic number. Although the emulsion layer is only 5 microns thick and only a small part of its volume consists of AgBr, the contributions of Ag and Br are nevertheless so substantial that the layer absorbs nearly as strongly as the layers of fig. 1*a*, and even more strongly below 3.7 Å.

While the great penetrating power of X-radiation is one of the fundamentals of microradiography, it is also necessary, of course, that the different parts of an object should exhibit distinctly different degrees of transparency, that is to say the rays must be differentially absorbed by the object. In view of the extreme thinness of the specimens (10-100 microns) this means that μ must be high. From equation (2) it is immediately apparent that soft X-radiation is required to get a large value of μ, although the gain in contrast must be paid for by a longer exposure.

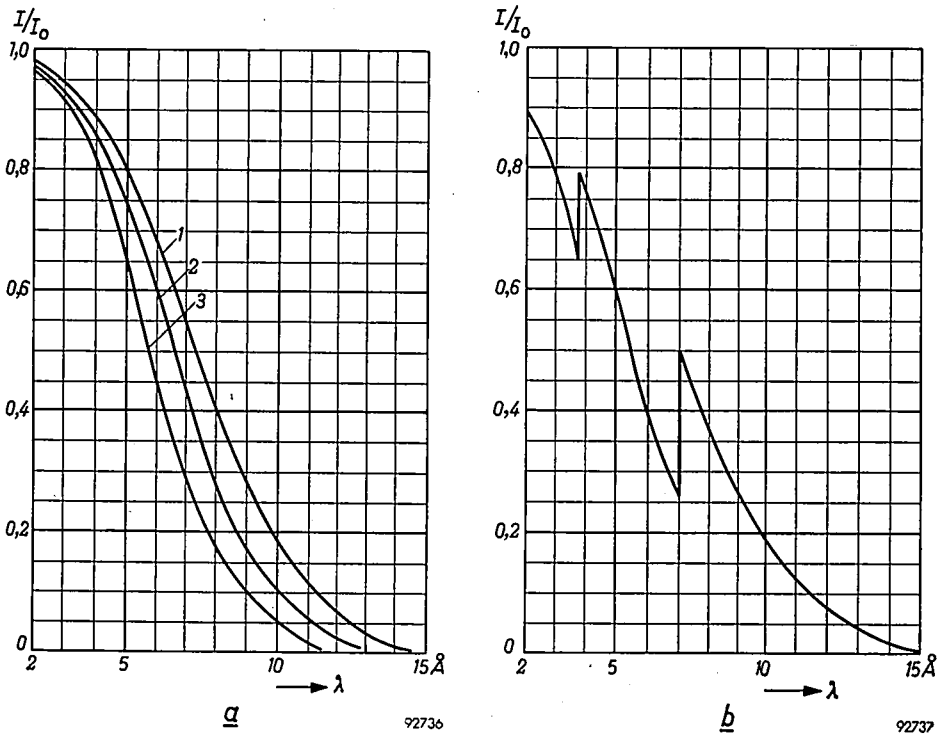


Fig. 1. *a*) Relation between the intensity of transmitted and incident X-radiation as a function of the wavelength, for three different absorbent layers:
1 animal tissue 10 microns thick, *2* layer of air 1 cm thick, and *3* beryllium foil 50 microns thick.
b) The same for the 5 μ thick emulsion layer of Kodak "Maximum Resolution" film. The discontinuities in the curve are the L absorption edges of bromine (left) and silver (right).

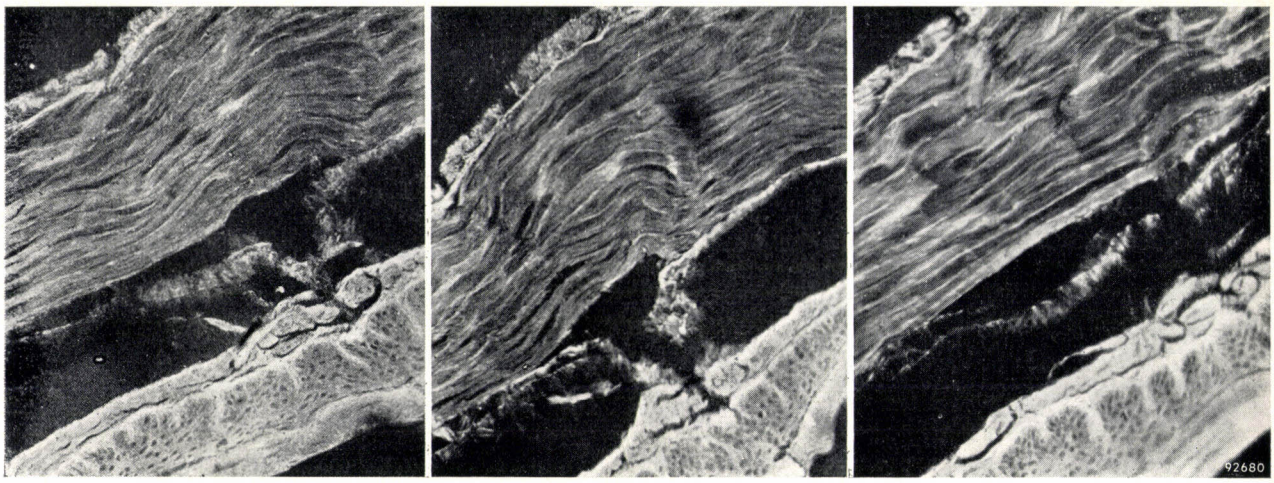


Fig. 2. Section of the oesophagus of a guinea pig (cavia). The light part is the mucosa (a glandular layer) on the inner oesophagal wall. The micrographs, from left to right, were taken at anode voltages of 5, 3 and 1.5 kV respectively, on Kodak "Spectroscopic" film 649-0. As can be seen, the contrast changes with the anode voltage. Magnification 130×.

It may be deduced that the film will show maximum contrast between two parts of an object with only slightly differing values of μ if the product μd is approximately equal to unity [4]). If we have a section of 20 microns thickness, i.e. 2×10^{-3} cm, then μ must be about 500 cm^{-1} in order that μd be equal to unity. For soft, animal tissue consisting of, say, 25% dry matter, this is the case when $\lambda = 11$ Å; for hard tissue (bone) the optimum wavelength with the same specimen thickness is 3 to 4 times shorter. Fortunately, however, fairly wide deviations from these optimum conditions cause no serious loss of contrast. The three micrographs in *fig. 2*, representing a cross-section of the oesophagus of a cavia (guinea pig), illustrate how the contrast changes as the X-radiation is made softer.

Projection microradiography and contact microradiography

A radiograph can be taken either by bringing the object and the film into close contact or by placing the film at some distance behind the object. In the first case the X-ray image is as sharp as it can be; in the second case there is some magnification, the X-ray image being larger than the object owing to projection (projection radiography [5])). The drawback of projection is that it gives rise to a penumbra (geometric unsharpness), the width of which increases with increasing distance between object and film. This phenomenon is illustrated in *fig. 3*.

It can be seen that the geometric unsharpness O_g may be expressed by the formula:

$$O_g = \frac{f\,b}{a-b}, \quad \ldots\ldots \quad (3a)$$

where f is the size of the tube focus and a and b are the distances between focus and film and object and film, respectively. By introducing the magnification V, equal to $a/(a-b)$, we may write

$$O_g = (V-1)f. \quad \ldots\ldots \quad (3b)$$

If object and film are in contact, O_g is very small, since in that case $b \approx 0$ or $V \approx 1$.

To determine the resolution obtainable, O_g should be compared with the linear dimensions of the *image*, which means that the linear dimensions of the

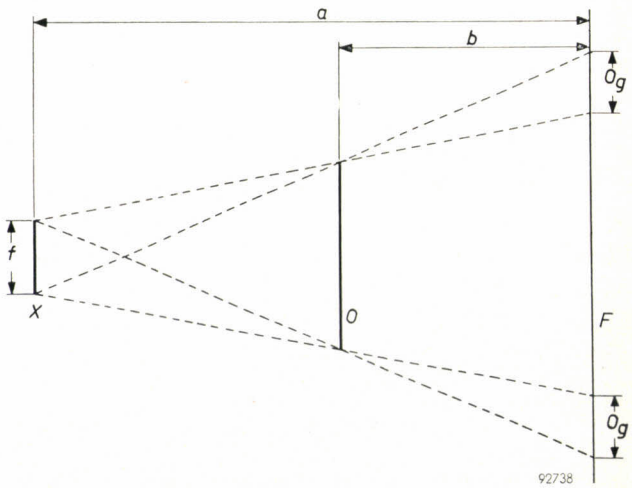


Fig. 3. The geometric unsharpness O_g of a microradiograph is determined by the dimensions of the focus (X) and the distances from film (F) to focus and from film to object (O). If the latter distance is very small (contact method), O_g is also very small.

[4]) A. Engström and R. C. Greulich, J. appl. Phys. **27**, 758-759, 1956.

[5]) See e.g. M. von Ardenne, Elektronen-Übermikroskopie, Springer, Berlin 1940, p. 72, and V. E. Cosslett and W. C. Nixon, Proc. Roy. Soc. **B 140**, 422-431, 1952.

object must be compared with O_g/V. With formula (3b) the resolution is then found to be

$$\frac{O_g}{V} = \frac{V-1}{V} f. \quad \ldots\ldots \quad (4)$$

This, then, is the smallest spacing there can be between two object details without their penumbrae overlapping, i.e. the smallest dimension a detail may have in order for its umbra still to reach the film. Expression (4) becomes very small when V approaches 1. At large V the resolution is practically equal to the diameter of the focus [6]).

To demonstrate this we shall take a numerical example. If the thickness of the object is 10 microns and that of the sensitive layer 5 microns, then b (formula 3a) in the case of contact is at the most 15 microns. Where the focus-film distance is 15 mm and the diameter of the focus 0.3 mm (300 microns), it follows from formula (3a) that O_g is at the most 0.3 micron. If, however, other conditions being equal, we take a projection radiograph with $b = 7.5$ mm, then $O_g = 300$ microns and, since $V = 2$, the smallest spacing separately distinguishable in the object is 150 microns.

Nevertheless, the above does not permit the conclusion that contact microradiography always produces better resolution than projection microradiography, for we have not yet taken the properties of the film into account. Owing to the fact that the sensitive layer is not a continuum but a suspension of light-sensitive AgBr grains in gelatine, the film itself has a limited resolving power. Let K be the smallest spacing that the film can resolve, then it is obvious that nothing more can be gained when $f(V-1)/V$ becomes smaller than K. With a very fine focus this limit is reached at a magnification considerably greater than 1. Thus, the spacing between two details which can be resolved, K/V, is thus in this case smaller than K, the threshold value in contact microradiography. If, for example, $f = K$, the optimum enlargement is $V = 2$ and the smallest distinguishable spacing is $K/2$.

The instrumental complications involved in the use of such an extremely fine focus, and the fairly low intensity of such an X-ray source, make the contact method preferable for many purposes, the more as it becomes possible to produce films with a smaller K. Moreover, the minimum tube voltage for projection microradiography is appreciably higher, owing to the necessity of using a transmission target [5]). In the following pages, therefore, we shall be concerned solely with contact microradiography.

The CMR 5 apparatus for contact microradiography

The necessity of using very soft X-rays has hitherto further complicated the technique of microradiography, since it was not possible to make sealed-off X-ray tubes with windows that would let through enough of the radiation required. Windows of sufficiently thin light-metal foil, such as beryllium and aluminium, were not vacuum-tight and the tubes employed had therefore to be continuously evacuated.

In recent years, however, it has proved possible to make vacuum-tight windows of beryllium foil only 50 microns thick. Because of the low atomic number of beryllium (4) the X-ray absorption in these windows is sufficiently small for the purposes of microradiography (see fig. 1*a*, curve *3*). An X-ray tube for microradiography equipped with such a window, and intended for anode voltages from 1.5 to 5 kV, is shown in *fig. 4*. A photograph of this tube is shown in *fig. 5*. The dimensions of the focus are 0.3×0.3 mm, and the focus-to-window distance is 11 mm. The fine focal spot is obtained by electrostatically focussing the electron beam, use being made of a cathode construction which has been

[6]) In accordance with the usage in microscopy, the term "resolution" denotes here the spacing between two details or objects that can just be separately distinguished.

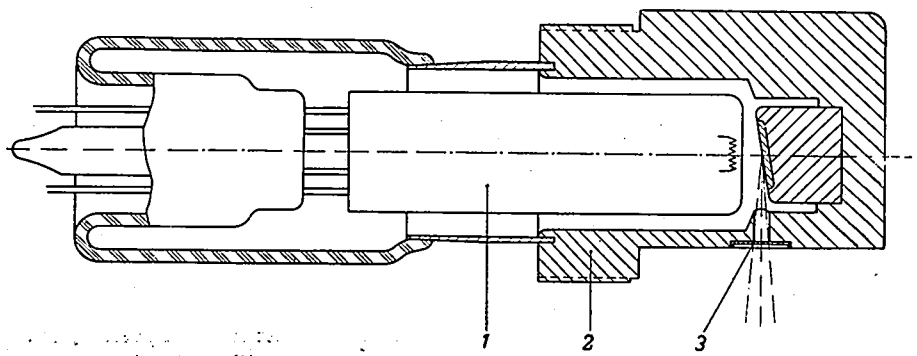


Fig. 4. Cross-section of X-ray tube for microradiography between 1.5 and 5 kV. *1* cathode with tungsten filament; *2* anode with tungsten target; *3* beryllium window 50 microns thick. The tube is 8 cm long. The (effective) size of the focal spot is 0.3×0.3 mm.

employed for some time now in fine-focus tubes for X-ray diagnostic technique [7]).

Fig. 6 shows a photograph of the CMR 5 apparatus in which the X-ray tube described is incorporated. The tube is mounted in the protruding section on the right, the window being underneath.

The apparatus contains the power supply circuit for the tube, the anode voltage being continuously variable. There is also a control for varying the tube current, so as to prevent over-loading of the tube (maximum power 10 W). The filament current is stabilized by means of a current stabilizer tube. The anode voltage and tube current can be read from meters on the front panel. After the controls have been set at the required values, the anode voltage can be switched off (in order, for instance, to load and mount the film-holder) without at the same time switching off the filament current, although the latter is then automatically somewhat lowered with a view to prolonging the life of the tube. In this way, steady operating conditions are rapidly reached when the

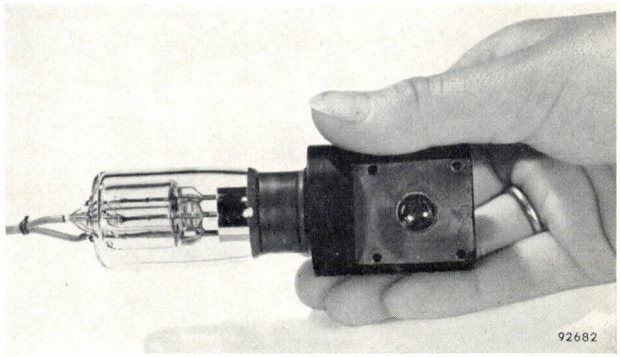


Fig. 5. Photograph of the X-ray tube shown in fig. 4.

[7]) See G. C. E. Burger, B. Combée and J. H. van der Tuuk, Philips tech. Rev. **8**, 327, 1946.

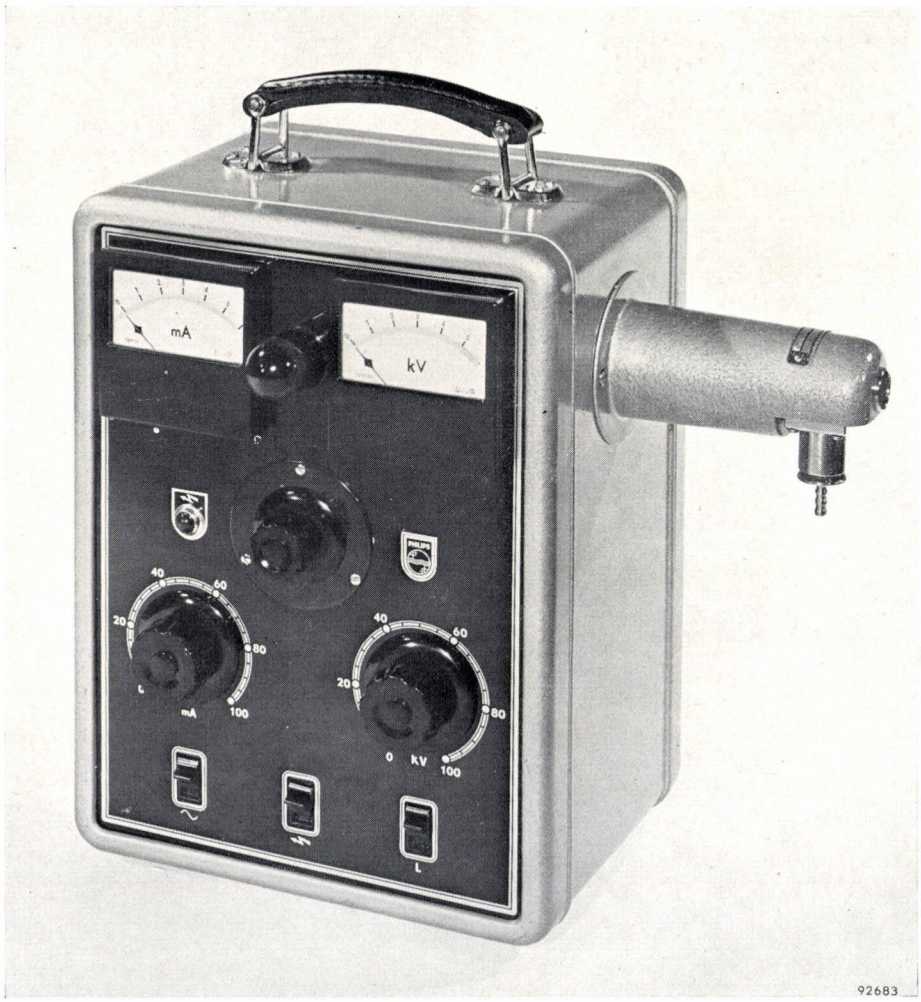


Fig. 6. The CMR 5 apparatus for contact microradiography. On the front panel can be seen the meters and controls for tube current (max. 5 mA) and anode voltage (max. 5 kV). The knob in the middle is used for fine adjustment of the current. The pilot lamp on the left is lit when the H.T. is on. The dial illumination lamp between the meters is switched on with the switch marked *L*. The other two switches are for mains and H.T. The X-ray tube is mounted inside the projecting section on the right of the case. The film-holder is inserted in the cylindrical container underneath the tube (see fig. 7).

high tension is switched on, so that it is not usually necessary to take warming-up into account when deciding exposure times.

The film-holder is mounted in the vertical cylindrical container under the X-ray tube. *Fig. 7* illustrates how this container is fixed to the X-ray

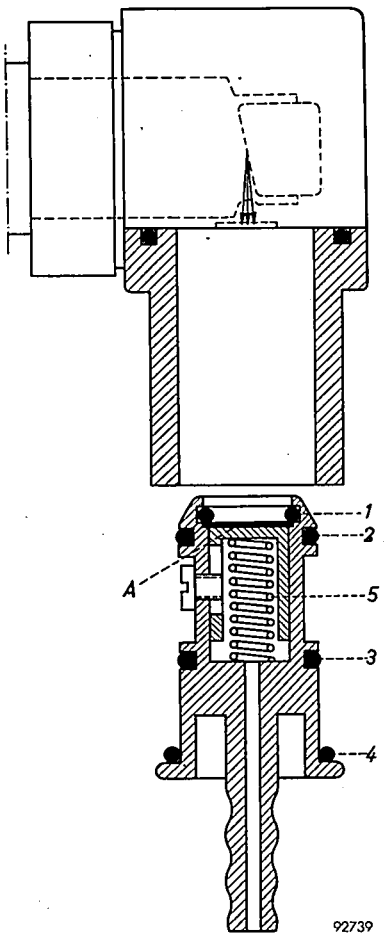


Fig. 7. The target end of the X-ray tube, showing attached cylinder into which the film-holder is inserted. Underneath is the film-holder itself, loaded with film, specimen and specimen carrier. The pressure plate is in the form of a plunger which, under the action of a spring *5*, keeps the film, specimen and specimen carrier (*A*) pressed against a rubber ring *1*. The other rubber rings (*2*, *3* and *4*) ensure a good seal, making it possible to evacuate the film-holder by attaching a vacuum line (a water-jet pump provides sufficient vacuum) to the nozzle provided for the purpose.

tube and shows a cross-section of the film-holder. In effect the film-holder is simply a cylinder with a spring-loaded plunger inside it; the plunger ensures that film, specimen and specimen base are kept pressed against the rubber ring *1*.

As can be seen in figs 6 and 7, the film-holder is provided with a nozzle, to which a vacuum-hose (connected, say, to a water-jet pump) can be fixed for the purpose of partially evacuating the film-holder and so reducing practically to zero the absorption in air which, as fig. 1*a* shows, cannot be left out of account.

For exposures longer than about 15 minutes, it is desirable to cool the X-ray tube; this can be done by connecting a fan to a point at the rear of the apparatus, the air then being drawn in through perforations in the protective shield around the end of the X-ray tube. This keeps down the temperature of film and object and prevents unsharpness arising from the mutual displacement of object and film due to thermal expansion.

Quality of the X-radiation

When applying the laws governing the absorption of X-rays in matter it should be borne in mind that the radiation generated in the apparatus and used for making the micrographs has a continuous spectrum. Ignoring the characteristic radiation also emitted, we can express the form of this spectrum by the equation [8]):

$$I(\lambda) \propto \frac{1}{\lambda^2}\left(\frac{1}{\lambda_0}-\frac{1}{\lambda}\right), \quad \ldots \quad (5)$$

where $I(\lambda)$ is the intensity of the radiation at a wavelength λ, and λ_0 is the shortest wavelength of the spectrum. Between the latter wavelength (in Å) and the anode voltage of the X-ray tube (in kV) there exists the relation:

$$\lambda_0 = \frac{12.39}{V}. \quad \ldots \quad (6)$$

Curve *1* in *fig. 8* shows the form of the spectrum after the radiation has passed through the beryllium window, the anode voltage being $V = 3$ kV. The minimum wavelength is 4.11 Å, but the maximum intensity occurs at 6 to 7 Å. It appears, then, that V must be made about $1^1/_2$ times larger than corresponds to the wavelength demanded by considerations of contrast and exposure time. The other curves in this figure show how the intensity is diminished after the rays have passed through a layer of tissue 10 microns thick and through the same layer plus the 5 micron thick emulsion of a Kodak "Maximum Resolution" film.

The film; magnification of the micrographs

Since there is very little geometric unsharpness in contact microradiography, provided the objects are sufficiently thin, the resolution obtainable is determined by the resolution of the film. The films most suitable for the purpose are Gevaert "Lippmann", Agfa "Mikrat", Eastman-Kodak "649-0" and Kodak "Maximum Resolution". These films have a resolving

[8]) H. A. Kramers, Phil. Mag. (6) **46**, 836-871, 1923. R. H. de Waard, Acta radiol. **28**, 37-48, 1947.

power of approximately 1000 lines per millimetre. Compared with most other films this figure is exceptionally high, but for microradiography an even higher figure would be desirable. The AgBr grains of the last mentioned film, for example, which have a diameter of about 0.045 micron, fill only a fraction of the total volume; there are some 200 grains per cubic micron and therefore about 6 per linear micron. Moreover, the grains are distributed not in perfect uniformity but at random. Obviously, in the image formed on such an emulsion it will be impossible to distinguish between details that are very much smaller than one micron, especially since a developed film, when viewed through a microscope, reveals a mottled structure of silver "grains" which consist of aggregates of the silver particles produced from the individual AgBr crystals during photographic development. It is this "graininess" that determines the resolution which the emulsion will give, it being understood, of course, that the resolution can never be better than the grain density of the original AgBr crystals permits.

In practice one can reckon on a resolution of 0.5 to 1 micron, which is almost as good as that obtainable visually with optical microscopes (0.3 micron). The micrographs can therefore be examined under a normal microscope with a magnification of, for example, 500×.

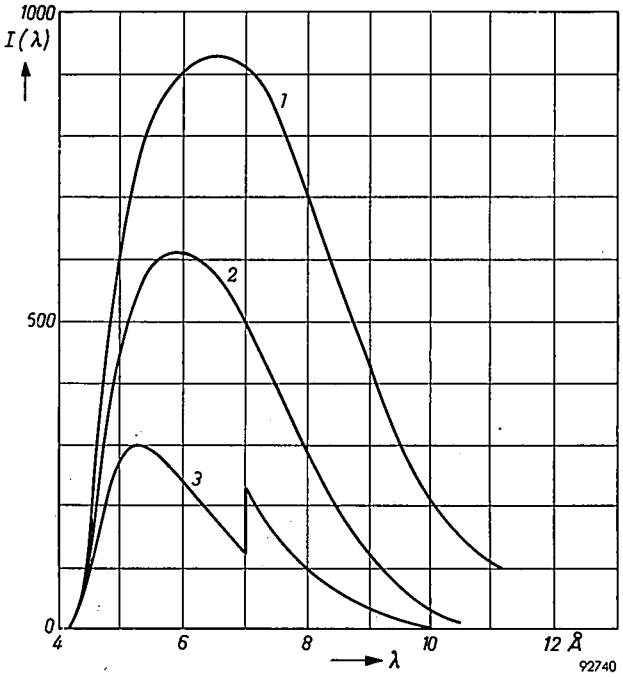


Fig. 8. Curve *1* shows the intensity distribution of the continuous spectrum of X-radiation generated in the CRM 5 apparatus at an anode voltage of 3 kV, and measured just outside the beryllium window. Curve *2* represents the spectral distribution of this radiation after it has passed through a layer of animal tissue 10 microns thick, and curve *3* after it has in addition traversed a 5-micron layer of emulsion on Kodak "Maximum Resolution" film. The discontinuity in the latter curve is due to the L absorption edge of silver.

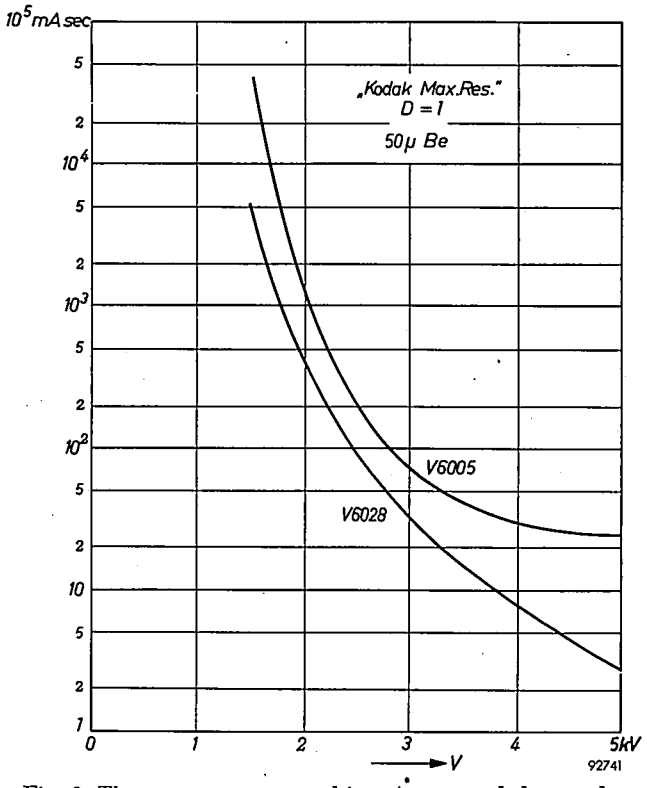


Fig. 9. The exposure, expressed in mA.sec, needed to produce a photographic density 1, as a function of anode voltage. The upper curve relates to standard Kodak "Maximum Resolution" film (emulsion V 6005) and the lower to an experimental Kodak emulsion V 6028. The curves show that the latter emulsion is at least 2× as sensitive as the other, and at 5 kV even 10× as sensitive.

The ideal emulsion for microradiography should have about the same absorption as the above, but should have a greater grain density and above all less graininess. Several film manufacturers have been attempting to produce emulsions with these properties, but their efforts have led rather to films of greater *sensitivity* — which is also an important point — than to films of higher resolving power.

This is illustrated in *figs 9, 10* and *11*, which relate to an experimental emulsion V 6028 by Kodak [9]). The experimental emulsion differs from the standard "Maximum Resolution" emulsion (type V 6005) in that it is only 1 micron thick instead of 5, and has 100 times more grains per unit volume; the size of the AgBr grains is the same. The curves in fig. 9 show what exposure, expressed as the product of tube current (mA) and time (seconds) is needed at a given anode voltage in order to produce a photographic density $D = 1$. The graph shows that the new emulsion is 2 to 10 times as sensitive as the old

[9]) We are indebted to Dr. R. W. Berriman and Dr. R. H. Herz (Research Laboratory Kodak Ltd.) for preparing this emulsion and placing it at our disposal.

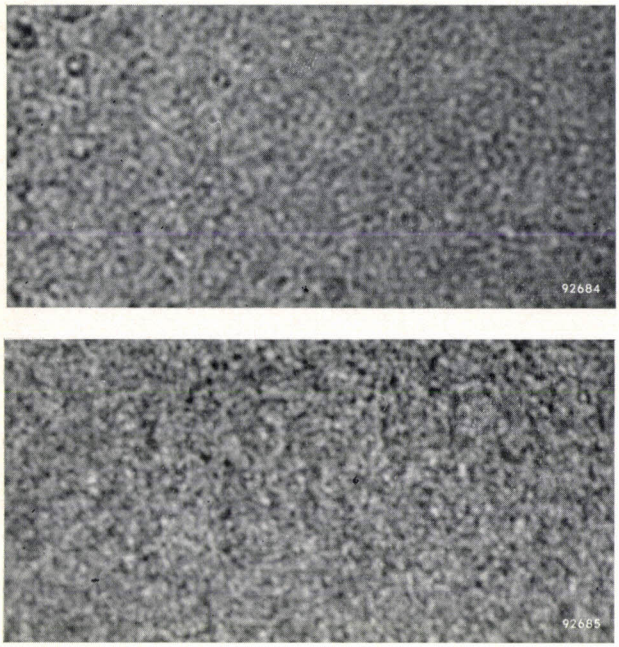


Fig. 10. Comparison of the graininess of standard emulsion V 6005 (above) and experimental emulsion V 6028, each with the same photographic density. Both exposures were made at an anode voltage of 3 kV. It can be seen that the new emulsion has about the same graininess as the old. Magnification 1200×.

one. Fig. 10 gives an impression of the graininess of the two emulsions with identical photographic density, and fig. 11 allows a comparison of the resolution in each case. The photographs show that the graininess of the new emulsion is neither more nor appreciably less pronounced than that of the other emulsion. It can be said, then, that the experimental emulsion V 6028 is much more sensitive than the "Maximum Resolution" type hitherto in use, without there being any decline in its resolving power.

It has been noticed that the resolving power of an emulsion is greater the softer the X-radiation employed. Experiments have shown that at anode voltages less than 1 kV it is possible to distinguish on one of the above-mentioned films details as small as 0.2 micron. In that case, magnification by an optical microscope is no longer adequate and an electron microscope must be used instead. We shall return to this subject presently.

Applications

Microradiography is used most for biological and medical research, particularly in the fields of histology and pathological anatomy. In these fields the technique can be put to the following uses to provide information unobtainable by optical microscopy.

a) Examination of opaque objects (e.g. bone tissue).

b) Examination of transparent objects, parts of which become visible because their chemical com-

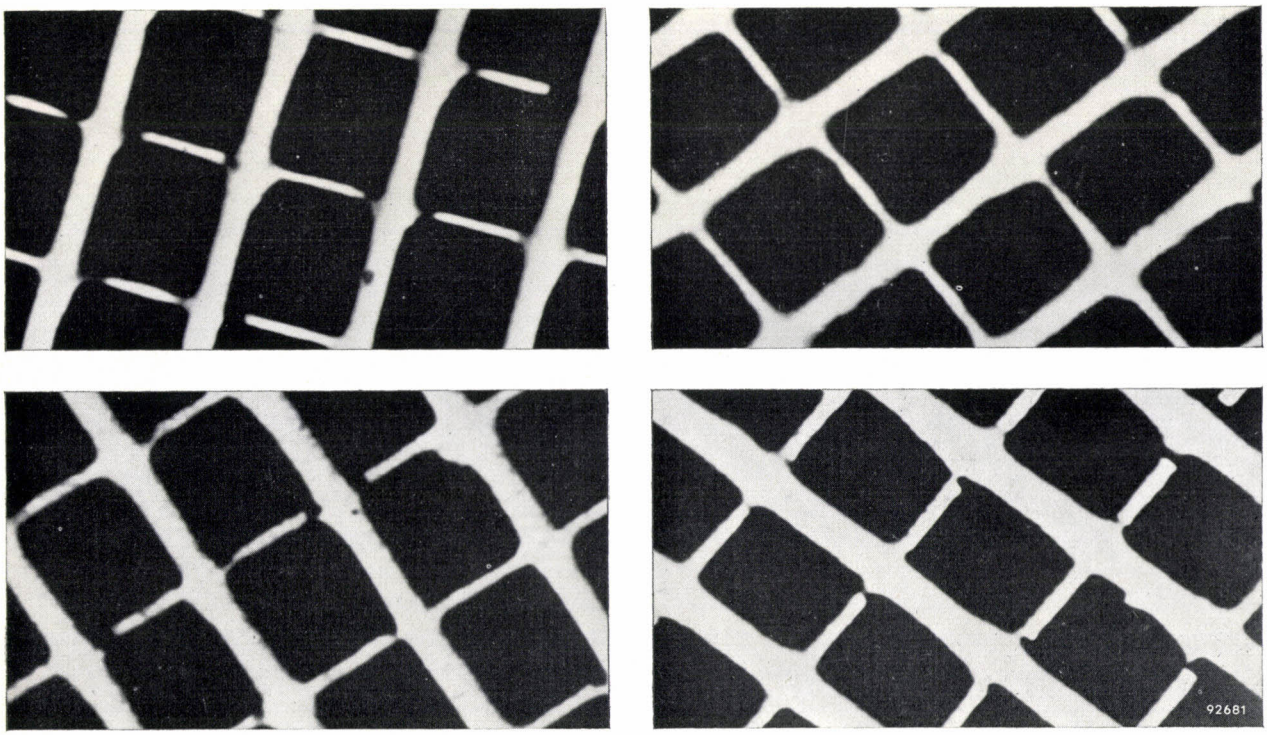


Fig. 11. Comparison of the resolution of emulsion V 6005 (above) and emulsion V 6028 (below). The two micrographs on the left were taken at an anode voltage of 3 kV, the two on the right at 1.5 kV. The object radiographed was a fine silver mesh, the thicker wires being 6 microns in diameter, the thinner 2 microns. It can be seen that the resolution of the experimental emulsion is roughly equal to that of the old; at 1.5 kV it is perhaps slightly better.

position differs substantially from that of their environment, or can be made visible by contrast media; this is especially important with regard to parts that are not easy to find with the aid of the dyes commonly used in optical microscopy.

c) Determination of the (dry) weight per unit area of an object, or detail, of known chemical composition, from the degree of X-ray absorption.

d) The performance of localized quantitative chemical analyses when the composition is known qualitatively.

In the following we shall briefly discuss, or exemplify, the four applications mentioned [10]). Before doing so, however, we shall give a synopsis of the technique employed in optical microscopy for preparing specimens of biological objects, and consider what parts of this technique can be adopted, omitted or modified for preparing specimens for microradiographic examination.

Preparation of specimens [11])

The procedure for preparing specimens for examination with an optical microscope is carried out in the following six stages:

1) *Fixation.* When a fragment is removed from a living object, that fragment slowly dies and thereby undergoes all manner of changes. To prevent such degeneration, the fragment is quickly immersed in a fluid which rapidly penetrates and kills the tissues and so fixes the fragment in as life-like a state as possible. Depending upon the object, the fixing fluids employed include alcohol (96%), acetic acid, formalin and osmium tetroxide (OsO_4), sometimes referred to as osmic acid.

2) *Embedding.* In order to be able to cut the fixed fragment into sufficiently thin sections, it is necessary, as a rule, to embed it in (i.e. thoroughly impregnate it with) a hardener such as paraffin wax. For this purpose the fragment is first dehydrated by immersing it in increasingly concentrated solutions of alcohol in water (in about ten stages). After that it is immersed (in three stages) in solutions of alcohol and xylol, the xylol content being increased at each stage, and finally, via a solution of paraffin wax in xylol, embedded in pure paraffin wax.

3) *Sectioning.* The embedded fragment is cut into thin sections with a microtome.

4) *Attaching specimen to slide.* This is done by applying first a drop of protein-glycerine to the slide, and then a drop of distilled water, taking care that it spreads well over the glass surface. The specimen is now placed on the slide and gently warmed (without melting the paraffin wax) until it lies flat. Finally, the water is drained off and the specimen left to dry in air, after which it is treated with xylol to remove the paraffin wax.

5) *Staining.* After the paraffin wax is removed, the procedure mentioned under 2) is reversed to return to the hydrated state. It is then possible to stain the specimen. The choice of stain depends on what detail one wishes to see. Among the frequently used stains are such synthetic dyes as methyl-blue, fuchsin, eosin (for staining the plasma), and natural dyes such as hematoxylin (for staining cell nuclei).

6) *Mounting.* The specimen is finally rinsed with alcohol and xylol and mounted in Canada balsam after which a cover-glass is placed over it.

For radiographic examinations the procedure is the same as described in 1) to 3) above. After that, however, the specimen is not mounted on a slide but on a formvar film stretched across a metal ring. Since formvar is resistant to xylol, the paraffin wax can be removed by the normal process [12]). No staining is involved, although a contrast medium may be applied where necessary. The mounting procedure under 6) is, of course, also irrelevant here.

For some purposes where the requirements are not so strict, the preparation of the specimens can be simplified. For example, the fixation and embedding procedures can be replaced by freezing, a special microtome then being used for sectioning. In this way the time spent on specimen preparation can be reduced from 1 or 2 days to 1 or 2 hours.

Examination of opaque objects

The advantage of contact microradiography over optical microscopy is obvious in the examination of opaque objects. *Fig. 12* shows a micrograph of bone tissue. Around the blood vessel the density is appreciably greater (i.e. there is less X-ray absorption) than at some distance away.

[10]) The latter two fields are dealt with comprehensively in an article by A. Engström, Historadiography, in Physical techniques in biological research Vol. 3, Acad. Press, New York 1956.

[11]) For a full treatment of this subject, see e.g. P. Metzner-A. Zimmermann, Das Mikroskop, F. Deuticke, Leipzig and Vienna 1928.

[12]) The formvar film (polyvinylformaldehyde) is only a few tenths of a micron thick and absorbs X-radiation only very slightly.

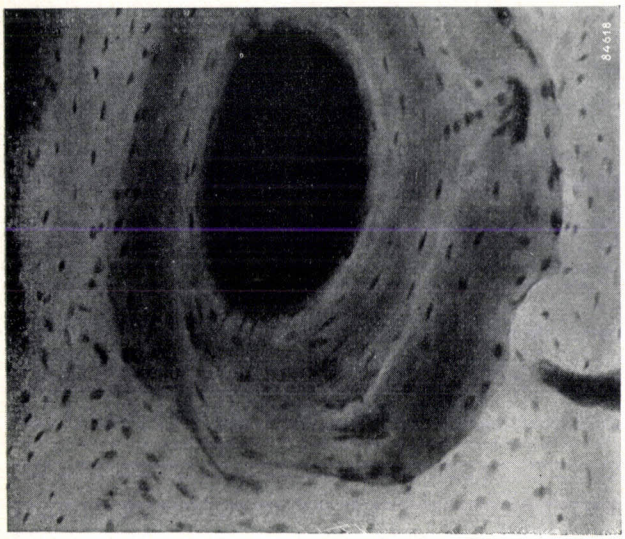


Fig. 12. Bone tissue around a blood vessel. To take such a microradiograph it is not necessary to de-calcify the bone, as it is in optical microscopy; hence it is possible to obtain information on the distribution of mineral salts in the bone. (5 kV, 3 mA, "Maximum Resolution" film; exposure 25 min.)

With micrographs of this kind it is possible to study, owing to the high absorption coefficient of the element, the deposition of calcium in bone tissue. It was in this way that Engström discovered the nature of the bone disease "osteogenesis imperfecta": microradiographic examination showed that there was incomplete calcification of the bone at various places and sometimes even rupture of the bone tissue.

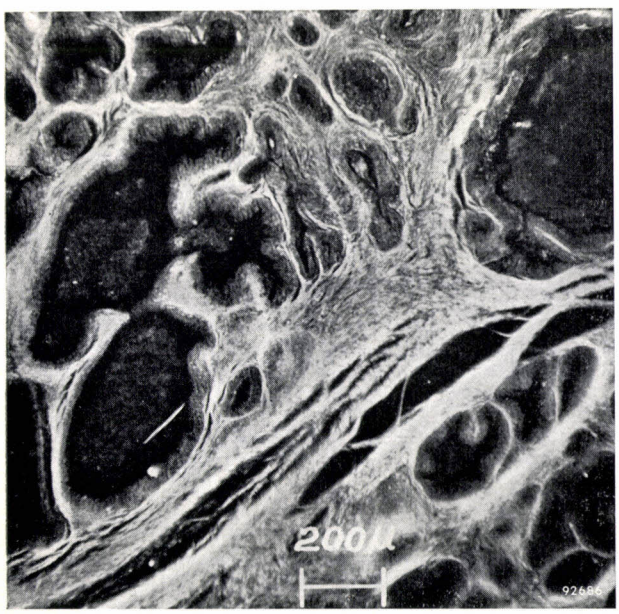


Fig. 13. Section (10 microns thick) of a hypertrophic prostate. The difference in the concentration of mineral salts in glandular and connective tissue is clearly visible. (3 kV, 3 mA, Eastman-Kodak "649-0" film; exposure 45 sec.) Explanatory note: the connective tissue appears very light, the glandular tissue grey; the large, very dark parts are due to the lumen, i.e. a space filled with fluid, generally serving for transport.

Examination of transparent objects

Just as it is possible to ascertain differences in chemical composition on optically opaque objects, it is also possible to do so on transparent objects. A good example is given in *fig. 13*. Without a complete chemical analysis the micrograph shows how the mineral salts have accumulated in the connective and not in the glandular tissue.

When there is very little or no contrast on the micrograph between various parts of an object in the natural state, contrast can sometimes be artificially produced with the aid of contrast media. For this purpose it is necessary to use substances containing heavy elements and which have the property of accumulating at very specific places instead of penetrating the object uniformly. For example, $AgNO_3$ is often used for showing up nerve fibres, Os compounds for fatty tissue and, for botanical specimens, I dissolved in KI solution, is used for staining starch.

An elegant application of this method is to follow what takes place during fixation with OsO_4. The latter — which is widely used in electron microscopy — is a slow process and brings about changes (naturally undesirable) in the specimen. In the long run, for instance, it causes the disappearance of the transverse bands of striated muscular tissue. A better insight into this process can be obtained by microradiographic examination [13]).

Determination of weight per unit area of biological objects

There are various methods of determining an object's weight per unit area. The most common is to radiograph, together with the specimen, a comparison object [14]), which may, for example, be a "step wedge" of collodion strips, laid upon the other, as shown in *fig. 14*. For this method it is necessary to have a not too small surface which is *homogeneously* irradiated by the X-rays. Not long ago a method was reported in which a comparison object is not required [15]). Several exposures of varying duration are made on a single film, and these give the relation between photographic density and exposure. Although a means of comparison is needed for obtaining the relation by which the mass can be derived from the transmission, all that is needed for this purpose is a layer of air of known thickness, temperature and pressure.

---

[13]) A. Recourt, Report of the Symposium on X-ray microscopy and microradiography, Cambridge 1956.

[14]) B. Lindström, Acta radiol., Suppl. **125**, 1955.

[15]) B. Combée and A. Engström, Biochim. biophys. Acta **14**, 432-434, 1954.

Quantitative chemical analysis

As already remarked, objects whose chemical composition is qualitatively known can be quantitatively analysed in their different details with the aid of microradiography. To determine the percentage content of a given element in the specimen, two micrographs are taken at wavelengths close to, but on different sides of, an absorption edge of that element. The percentage is derived from the ratio of the transmissions and the known variation of the absorption coefficient with the wavelength of the radiation [16]). For these measurements, then, a monochromatic X-ray beam is needed, which means that they cannot, without further measures, be carried out with the CMR 5 apparatus.

To sum up, it may be said that the CMR 5 apparatus constitutes a great advance towards the goal of making microradiography as simple a technique as optical microscopy. At a cost comparable to that of a good optical microscope, it has brought microradiography within the reach of many who have hitherto been deterred by the complicated equipment once required. The preparation of specimens takes less time than in optical microscopy, and represents no insuperable obstacle to the use of microradiography for routine pathological work. Moreover, it calls for no technique that was not already familiar in optical microscopy.

The salient features of the new apparatus are: 1) the use of a sealed-off X-ray tube, 2) small and simple H.T. generator, 3) compact construction, 4) very small focus, hence such small geometric unsharpness that the resolution is determined solely by the film (0.5 micron approx.). The apparatus can be used for all the purposes discussed in the foregoing, except for quantitative chemical analysis.

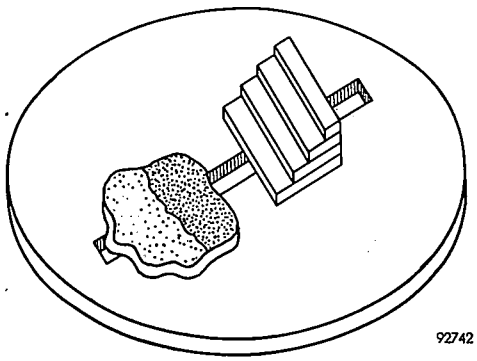


Fig. 14. Schematic representation of specimen carrier, showing specimen and collodium step-wedge side by side. From the known wieght per unit area of each step and the associated photographic density, a curve can be plotted with the aid of which the weight per unit area of a given part of the specimen can be found from the density of the corresponding part of the microradiograph.

[16]) See Part III of [14]).

Experiments with voltages lower than 1 kV

As we have seen, the resolving power of an emulsion is improved by the use of X-radiation of longer wavelength. It is interesting to see how this phenomenon continues at anode voltages still lower than the 1.5 kV limit with the CMR 5 apparatus. It has been found that at wavelengths corresponding to an anode voltage lower than 1000 V the smallest spacing that the types of films mentioned are capable of resolving is about 0.2 micron [17]). To derive full advantage from this high resolution, magnification by an optical microscope is no longer adequate and use must be made of an electron microscope.

For these experiments [18]) use was made of an X-ray tube of the type used in the CMR 5 apparatus, but without its window. Object and film were mounted in the normal way, except that the object was covered by an Al foil 0.1 to 0.2 micron thick to prevent light from the glowing cathode from reaching the film. The whole arrangement is shown schematically in *fig. 15*. In this way it is possible, without having to resort to extremely long exposures, to use anode voltages as low as 500 V. Naturally, this procedure meant that every time the film-holder was reloaded the tube had to be re-evacuated. *Fig. 16* shows two micrographs of a cavia oesophagus which were taken at anode voltages of 1000 and 850 V respectively (cf. fig. 2). *Figs 17* and *18* show the excellent resolution obtainable. Both represent a detail from fig. 16, the first magnified by an optical microscope, and the second by an electron microscope.

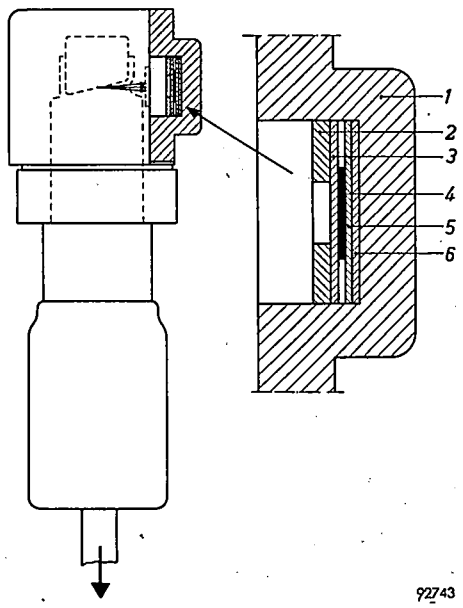


Fig. 15. X-ray tube and film-holder for contact microradiography at anode voltages lower than 1.5 kV. The tube is the same as that used in the CMR 5 apparatus, but the pump stem is not sealed off and the window is missing. In the film-holder (*1*) are mounted successively (seen from the target) the specimen mounting ring (*2*), a collodion film (*3*) carrying the specimen (*4*), on which an Al layer (*5*) is vapour-deposited to keep light from the cathode away from the film, and finally the film (*6*).

[17]) R. C. Greulich and A. Engström, Exptl. Cell Research **10**, 251-254, 1956.

[18]) See note [13]).

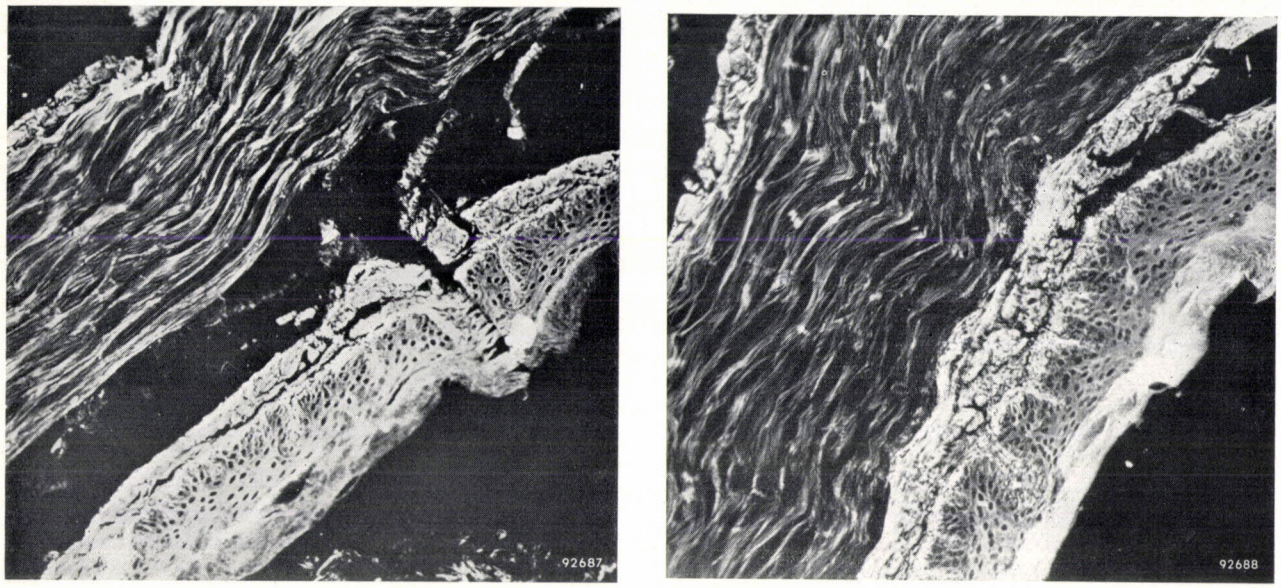


Fig. 16. Section of the oesophagus of a cavia, on the left taken at 1000 V, on the right at 850 V. It can be seen that the resolution is better than in the micrograph taken at 1.5 kV, shown in fig. 2. Magnification by optical microscope (130×).

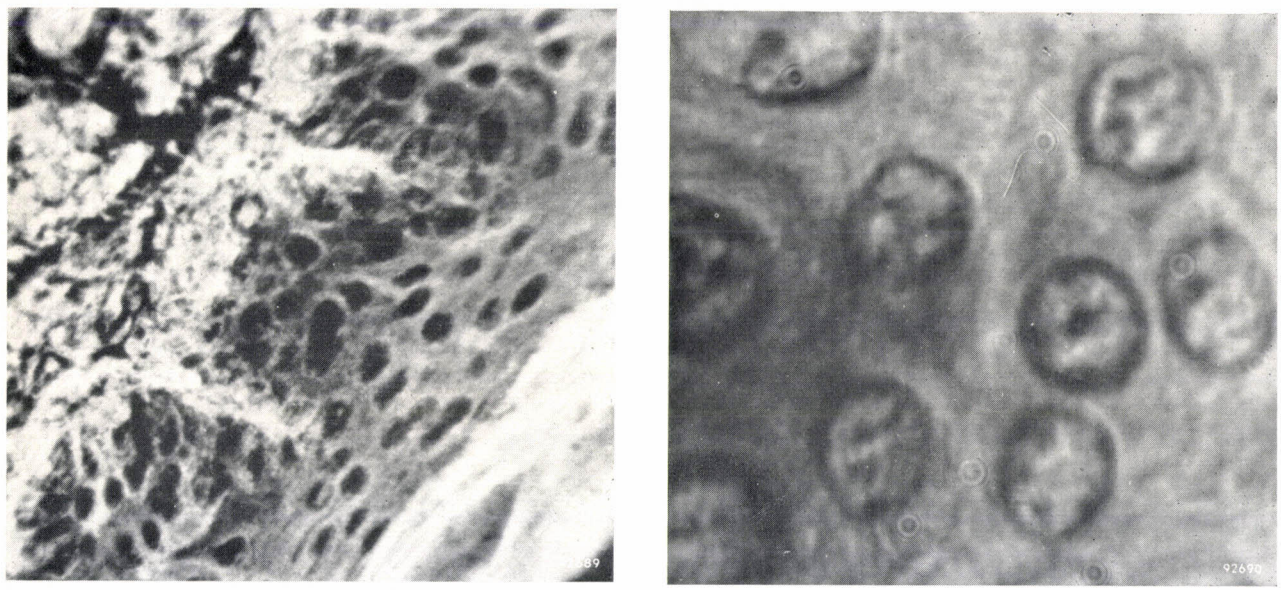


Fig. 17. Detail of 850 V micrograph in fig. 16. Magnification by optical microscope (600×).

Fig. 18. Detail of 850 V micrograph in fig. 16. Magnification by electron microscope (2000×).

Magnification by an electron microscope introduces a new difficulty, since a film is too thick to act as an object for an electron microscope. As far as possible, all superfluous material must be removed (the celluloid film and the gelatine of the emulsion), leaving finally only the developed silver grains on a formvar film. This can be done, though the procedure — partly based on a technique of Comer and Skipper [19]) — is rather complicated.

[19]) J. J. Comer and S. J. Skipper, Science **119**, 441-442, 1954.

Summary. To make contrasty radiographs of thin sections of biological matter (e.g. 10 microns thick), it is necessary to use soft X-radiation. Such soft radiation cannot leave the X-ray tube unless the latter can be fitted with an exceptionally thin window of a material having a low atomic number. Until recently such windows could not be made vacuum-tight, and therefore the tube had to be continuously evacuated on a pump during the taking of an exposure. In latter years, however, it has proved possible to make a sealed-off X-ray tube, suitable for anode voltages from 1.5 to 5 kV, fitted with a beryllium window only 50 microns thick. A portable apparatus for contact microradiography (CMR 5), equipped with this tube, has now been developed. The apparatus contains a small H.T. generator for the anode voltage which can be varied continuously up to 5 kV. The tube current can also be regulated (max. 5 mA). The filament current is stabilized. When the H.T. is switched off, the filament voltage is automatically lowered to conserve the tube. Provisions are made for cooling the tube during long exposures and for evacuating the film-holder (e.g. with a water-jet pump) to prevent absorption of the X-rays in air. Since the focal spot is very small (0.3×0.3 mm) and film and object are in close contact, geometric unsharpness is negligible, and therefore the resolution is entirely determined by the resolution of the film. It is as good as 0.5 to 1 micron, which approaches that obtainable visually with an optical microscope (0.3 micron).

This means that an ordinary optical microscope is quite capable of providing the magnification required. Experiments are being carried out by several film manufacturers with a view to producing films with a better resolving power.

The procedure of preparing specimens is similar to that used in optical microscopy, but can be considerably simplified, reducing the time needed from 1-2 days to 1-2 hours. The particular value of microradiography is that it makes it possible to examine details in otherwise opaque objects (e.g. bone) and to gain an insight into more than merely the geometry of an object — for example the distribution of mineral salts in a tissue or the weight per unit area of particular details. If a monochromatic X-ray beam is available, the technique can also be employed for quantitative chemical analysis of separate details of an object. Examples of such problems are discussed with reference to the practice of histology and pathological anatomy.

Finally, some experiments are described in which no window is interposed between X-ray focus and object and the generating voltage is lower than 1 kV. With the soft X-radiation so produced, the films normally used for microradiography show a resolution as good as 0.2 micron. To take full advantage of this, the micrographs must be magnified with an electron microscope.

A MIRROR-CONDENSER LAMP FOR 8-mm PROJECTORS

621.326.732:778.55

To obtain a high luminous flux on the screen from 8-mm film projectors, designers tended until recently to use high-power lamps rather than to look for ways of improving the condenser optical system. The use of high-power lamps entails penalties in the size, weight and price of the projectors.

Closer examination of this problem in the Philips Laboratory at Eindhoven led to the conclusion that it must be possible to achieve the same result with a low-power lamp, by abandoning the conventional condenser optical system in favour of a small incandescent lamp having a partly silvered bulb, the latter being so designed as to serve as a mirror condenser. Experiments [1]) have confirmed this conclusion.

The lamp in question, which consumes only 50 watts (8 V, 6.25 A) is shown in *fig. 1*. *Fig. 2* represents a horizontal cross-section at filament level. The rear portion of the bulb has the form of an ellipsoid of revolution and is externally coated with a metal reflecting layer. The filament, a horizontal single spiral, is mounted at one focus of the ellipsoid, and a magnified image of the filament is formed at the other focus, which is situated about 12 mm beyond the front wall of the bulb. The film gate — 4.45 × 3.35 mm in 8-mm projectors — can now, without further optical components, be set up at about the position of the second focal point. It thus receives sufficiently uniform illumination from a beam of light having an aperture angle large enough for filling a projection lens of conventional speed (relative aperture f: 1.5) [2]). The front portion of the bulb is spherical in form and is silvered over an annular region as shown in fig. 1. By this means, some of the light emitted to the right (fig. 2), which would otherwise be lost, is reflected back on to the ellipsoidal mirror. Owing to the low power dissipation of the lamp, no forced cooling is required.

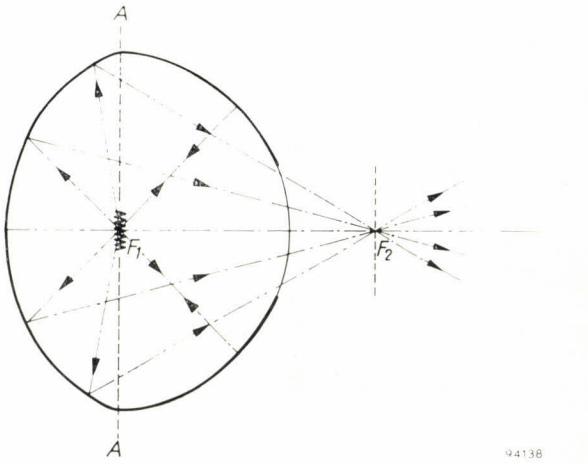


Fig. 2. Horizontal cross-section through mirror-condenser lamp at filament level. The section to the left of the line AA is part of an ellipse with foci F_1 and F_2. The section to the right of AA forms part of a sphere.

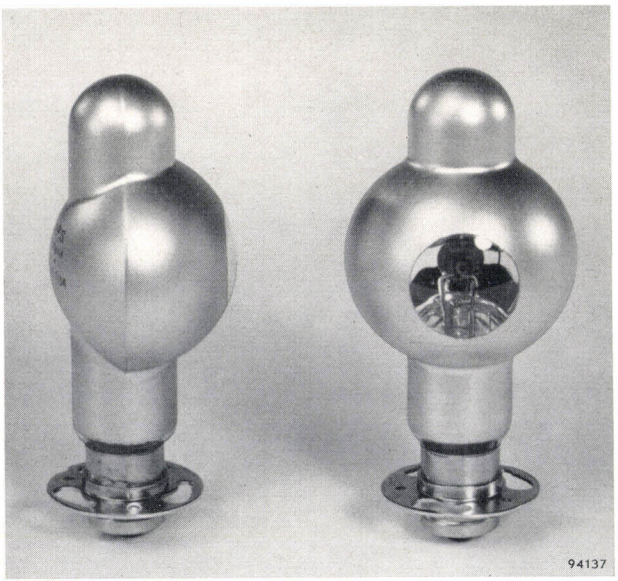


Fig. 1. Mirror-condenser lamp for 8-mm projectors. The total height of the lamp is 94 mm.

[1]) By F. L. van Weenen, who also posed the problem.

[2]) Elliptical mirror condensers have been used for other purposes (standard-film projection, airfield and stage lighting). See e.g. R. Sewig, Handbuch der Lichttechnik, part II, p. 710, Springer, Berlin 1938, and F. W. J. Schweigman, thesis, Groningen 1946.

Before discussing the details of this system, we shall briefly consider in general terms the two methods of arranging a condenser system to give intense uniform illumination of a projection aperture — and hence of a projection screen.

With the first arrangement (*fig. 3*), which is used as a rule for the projection of lantern slides, the condenser throws an image of the light-source on the projection lens and the diapositive is situated very close behind the condenser. This produces highly uniform illumination of the diapositive. If

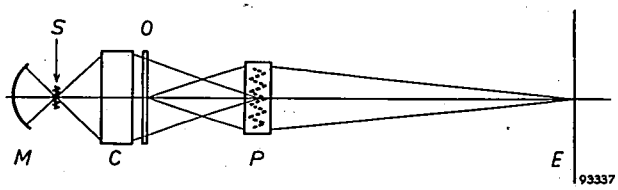


Fig. 3. Path of rays in a lantern-slide projector. The image of light-source S is formed in the projection lens P. The lantern slide O is immediately behind the condenser C and is thus uniformly illuminated. The auxiliary mirror M causes a somewhat larger fraction of the luminous flux emitted by the lamp to be usefully employed, but plays no part in the image formation. The projection lens projects the image of O on to the screen E.

this (physically ideal) system were to be used for the projection of very small pictures, and in particular 8-mm cine film, it would be necessary — assuming, that is, that loss of light at the film gate has to be kept to a minimum — to construct condenser lenses and light-sources of corresponding dimensions. Apart from the question of whether this would be technically feasible, it would still not be a satisfactory solution since condenser and film would be damaged by the heat generated by the light-source placed in such close proximity to them [3]).

The other arrangement, which is normally employed in film projectors, is illustrated in *fig. 4*. In this case the light-source is imaged in the film gate. Here, too, very little is lost of the light collected by the condenser, but the illumination of the film gate

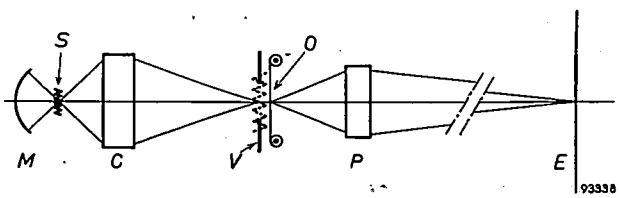


Fig. 4. Path of rays in a film projector. The image of the light-source S is formed near the film gate V by the condenser C. The auxiliary mirror M has the same function as in fig. 3. P projects the image of film O on to E.

[3]) In the projection of 35-mm film this drawback can be overcome by using a mercury-vapour lamp as light source. See Philips tech. Rev. 4, 2-8, 1939.

is sufficiently uniform only if — assuming ideal image formation — the light-source has a surface of approximately constant brightness. Moreover, in order to limit light losses, this surface should be practically identical in form with the film gate.

With both these systems it is possible in principle to replace the condenser lenses by a mirror. It is even advantageous to do so, having regard to the solid angle subtended by the mirror at the filament. Mirrors with an aperture angle of 180° can readily be constructed, whereas not much more than 90° is possible with lenses, without incurring heavy costs. What is more, a mirror is, of course, free of chromatic aberration. In practice, however, the use of mirror condensers has so far been restricted to 35-mm projectors (with a carbon arc as light-source), the reason being that these condensers function satisfactorily only when their dimensions are large with respect to the light-source. If an incandescent lamp is used in conjunction with a separate mirror, the critical dimensions to be taken into account are rather those of the bulb than those of the filament: where the mirror employed is not large compared with the bulb a considerable fraction of the reflected light again passes once or twice through the bulb wall and undergoes, by refraction, changes in direction which distort the formation of the image. That, above all, is why it has hitherto not been practicable to use a mirror condenser in 8-mm projectors. Another drawback attaching to mirror condensers is that, in the long run, the reflecting surface becomes tarnished.

It is immediately evident from what has been said that these disadvantages do not apply to a lamp having the mirror directly coated on the bulb; tarnishing is no longer a problem and moreover the refraction of the light in the bulb is now of neglibible significance.

When the mirror-condenser lamp is used in conjunction with a projection lens having the relative aperture often used for 8-mm projection ($f:1.5$) the luminous flux incident on the screen is about 100 lumens. Suppose now we set up an incandescent lamp having the same filament luminance, together with an auxiliary mirror (see figs 3 and 4) and a normal condenser lens subtending an angle of, say, 90° at the filament. With the same projection lens as used above (this implies that the light from the condenser must converge at the same angle as before) and assuming the same overall reflection losses of 50% as estimated for the previous case, the luminous flux incident on the screen is the same as before (100 lumens). However, a larger filament will have been used: this follows from the

conditions that the light must just fill the film gate and converge at the same angle as before [4]). One may also say that, since a smaller fraction of the total light emitted is caught by the condenser lens and the filament has the same luminance as before, the same light flux through the system can be obtained only by using a larger luminous surface — which means a higher power. In the case cited (condenser subtending 90° at filament) the 100 lumens on the screen represents only $2\frac{1}{2}$% of the total light emitted by the filament (see *Table I*). In the case of the mirror condenser lamp, however, the 100 lumens represents about $4\frac{1}{2}$% of the total flux from the filament. It follows that the powers of the two lamps which, in the two cases, produce the same luminous flux on the screen are approximately inversely proportional to the above percentages.

In reality the power of the mirror-condenser lamp is even less than might be supposed from a comparison with other lamps on the basis of the above percentages, for the reason that the filament is brighter. This is achieved by the use of a low supply voltage: the filament is relatively thick so that, for a given useful life, a higher temperature is permissible than in high-voltage lamps. Because of the low power, a projector equipped with the new lamp requires only a relatively light transformer, and this, together with the absence of a cooling system, helps to keep down the weight of the projector.

Table I. The fraction of the emitted luminous flux reaching the projection screen when using a condenser lens subtending an angle of 90° (dimensions of filament adapted accordingly) compared with that when using the mirror-condenser lamp, both in conjunction with a projection lens having a relative aperture f : 1.5. The figures are only approximate, but they demonstrate the considerable gain obtained with the new lamp, owing to the large angle subtended by the condensing mirror at the filament.

	Condenser lens	Mirror-condenser lamp
Luminous flux emitted by lamp . . .	100%	100%
Fraction of the above collected by condenser (angle subtended by condenser lens 90°, by mirror-condenser lamp 160°) *)	30%	56%
Shutter transmits, say, one half the above	15%	28%
Due to overlap of the filament image and the non-ideal geometry [4]), the film gate transmits about $\frac{1}{3}$ of the above **)	5%	9%
Overall reflection losses at lens surfaces, lamp wall and condensing mirror here expressed as a percentage of the light flux that would otherwise emerge from the projection lens: approx. 50% ***). Hence total incident on screen	$2\frac{1}{2}$%	$4\frac{1}{2}$%

*) The percentages for both lens and mirror condensers are higher than would be expected from the effective solid angles they subtend: this is due to the use of the auxiliary spherical mirror.

**) Only when the dimensions of the filament are appropriate to the geometry of the condenser, may the same fraction ($\frac{1}{3}$) be taken in both cases.

***) Probably less in the set-up employing the new lamp.

Since light-source and condenser form a single unit in the new lamp, a high degree of accuracy in dimensions must be achieved during fabrication. The lamp is provided with a type P15 pre-focussing base (see fig. 1). Each lamp is separately adjusted in an assembly jig, the filament being given a certain specified position with respect to the pre-focussing flange on the base; this having been done, the flange is soldered to the base [5]). Clearly, to obtain optimum results with the lamp, it is essential that the projector lampholder is also accurately positioned.

If we compare the properties that may be expected of an 8-mm projector equipped with the new lamp with those of the best existing projectors, we find that the number of lumens incident on the screen per kilogram weight of apparatus, and also the number of lumens per watt, are two to three times higher, an improvement, moreover, which is not obtained by using a very expensive lamp. On the contrary, the price of the lamp per lumen-hour is lower than that of other incandescent lamps for 8-mm projectors. The "price per lumen" of the whole projector will be even more favourable.

P. M. van ALPHEN,
M. BIERMAN.

[4]) See, for example, J. J. Kotte, A professional cine projector for 16-mm film, Philips tech. Rev. **16**, 158-171, 1954/55.

[5]) See Th. J. J. A. Manders, Philips tech. Rev. **8**, 72-81, 1946. This article also deals with other problems concerned with projection lamps.



THE FOUCAULT PENDULUM IN THE UNITED NATIONS BUILDING IN NEW YORK

by J. A. HARINGX and H. van SUCHTELEN.

525.36.002.2

Among the gifts presented by member states for embellishing the United Nations building in New York is a large Foucault pendulum from the Netherlands. The pendulum, which is suspended in the entrance hall leading to the main Assembly Hall, continuously demonstrates by its movement that the earth is in rotation.

The article below briefly describes the suspension and driving systems of this pendulum. The design was based on the requirement that the pendulum should function well for a very long time without supervision and maintenance.

On 7 December 1955, Mr. H. Luns, Dutch Foreign Minister, presented to the Chairman of the United Nations General Assembly a Foucault pendulum on behalf of the people of the Netherlands. The pendulum is suspended above the central column of the staircase in the entrance hall of the United Nations building in New York (see *fig. 1*). The pendulum was specially designed in the Research Laboratory of N.V. Philips' Gloeilampenfabrieken, the guiding consideration being that the pendulum should function uninterruptedly for many years without requiring supervision or maintenance.

A Foucault pendulum is in principle merely a weight (which we shall henceforth refer to as the "bob") suspended by a long wire, swinging in a vertical plane. Owing to the manner of suspension the plane of swing is not, however, fixed with respect to the earth but, in consequence of the rotation of the earth, turns about the vertical through the point of suspension. In the northern hemisphere the plane of swing deviates in the clockwise direction, in the southern hemisphere in the opposite direction. It can be shown [1]) that for an idealized case (very long pendulum, very small amplitude) the time T' in which the plane of swing rotates 360° is given by the formula:

$$T' = \frac{T}{\sin \varphi}, \quad \ldots\ldots \quad (1)$$

in which T is the period of revolution of the earth and φ is the geographic latitude. At the poles $T' = T$, while at the equator $T' = \infty$, i.e. there is no rotation of the plane. In New York ($\varphi = 40°45'$), time $T' = 36$ hours 50 minutes. It further appears from the theory that the angular velocity with which the plane rotates with respect to the meridian is constant.

[1]) See e.g. A. Sommerfeld, Vorlesungen über theoretische Physik, Part I, Dieterich Verl., Wiesbaden 1952.



Fig. 1. The entrance hall in the United Nations building in New York. The Foucault pendulum, a gift from the Netherlands, is mounted above the central column of the staircase.

In practical conditions, the time of rotation is found from the formula:

$$T'' = \frac{T}{\sin \varphi}\left(1 - \frac{3}{8}\frac{a^2}{l^2}\right), \quad . \quad . \quad . \quad . \quad (2)$$

in which a is the amplitude and l the length of the pendulum [2]). In the case of the pendulum in the United Nations building, l is about $17\frac{1}{2}$ m and a about 80 cm. This length gives a period of swing of approximately $8\frac{1}{2}$ sec. The weight of the pendulum bob is about 90 kg.

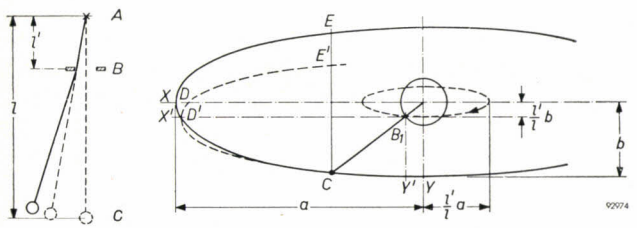
If we try to construct such a pendulum and repeat the celebrated experiment carried out by Foucault in 1851, a number of disturbing effects — due, for example, to imperfections in the rotational symmetry of the wire support — cause the pendulum, after some time, to execute an elliptical or even a circular motion. This had to be avoided with the present pendulum, which was required to swing continuously

[2]) Handbuch der Physik, Part V, p. 339, Springer, Berlin 1927.

without supervision. There were two other technical problems. In the first place there had to be the certainty that the wire would not break after some time of continuous operation, and in the second place a driving system was needed that would be able to provide the pendulum with sufficient energy to compensate for the loss of energy due to air resistance.

A simple method of overcoming the "ellipsing" problem has been described by Charron [3]). At a distance l' below the point of suspension A (see *fig. 2a*) he fixed a ring B having an internal diameter slightly larger than the thickness of the pendulum wire, leaving an annular spacing of d. As soon as the deflection of the bob exceeds the value ld/l', the wire touches the ring and the point of contact then functions as the "point of suspension". The consequence is that the minor axis of an elliptical orbit which may have been forming is rapidly diminished. With reference to fig. 2*b*, this can be roughly explained as follows.

Assuming that the wire, as long as it does not touch the ring, is straight, it will describe at the level of the ring an ellipse geometrically similar to the elliptical orbit of the bob reduced in the ratio l'/l. This is represented in the figure by the small broken ellipse; it is traversed, for example, in the direction of the arrow. The circle represents the limit set by the ring to the deflection of the wire. The moment the wire touches the ring, point B_1 functions as the point of suspension. Assuming that the velocity of the ball in the Y direction is small at that moment, the ball will thereupon proceed to describe the ellipse $CD'E'$, the axes of which lie along the lines B_1X' and B_1Y'. We see, then, that the amplitude



a b

Fig. 2. *a*) Suspension of pendulum according to Charron. The point of suspension is at A. The ring B, whose internal diameter is only slightly larger than the diameter of the wire, prevents the bob C from describing an ellipse.
b) Functioning of Charron's ring. When the pendulum describes an ellipse with semi-axes a and b, the wire touches the ring e.g. at point B_1. If there is sufficient friction between ring and wire, B_1 then functions for a while as the effective point of suspension and the bob describes the ellipse $CD'E'$ instead of the ellipse CDE, which it would describe in the absence of the ring. The amplitude in the Y direction is thus diminished in a half period by the length EE', which is equal to $2bl'/l$.

[3]) F. Charron, Bull. Soc. Astr. Fr. **45**, 457, 1931.

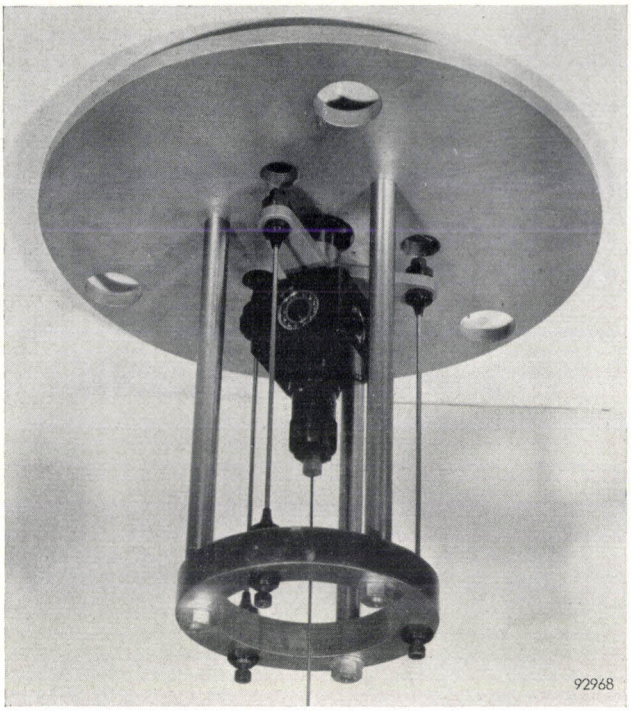


Fig. 3. The actual suspension system.

in the Y direction at the beginning of the following half swing has been reduced by the distance EE', which is approximately equal to $2bl'/l$. The relative reduction for each full swing amounts then to $\Delta b/b = 4l'/l$.

It has been tacitly assumed in the foregoing that while the bob describes the orbit $CD'E'$ the wire stays pressed against the ring at point B_1. As a rule the contact friction will not be sufficient for this to occur. In reality, therefore, $\Delta b/b$ is smaller than calculated here, but this does not alter the fact that the mounting of the ring is a very effective means of combating the elliptical motion.

Charron deduced that the use of this construction would slightly shorten the period of rotation and derived the following expression for the period:

$$T''' = \frac{T}{\sin \varphi}\left(1 - \frac{3}{8}\frac{a^2}{l^2} - \frac{4d}{\pi a}\right). \quad . \quad . \quad . \quad (3)$$

For the pendulum installed in New York a suspension was designed, wich, while equivalent to that of Charron, differs considerably in its construction. *Fig. 3* shows a photograph of the suspension and *fig. 4* a diagram indicating its construction. The mounting plate A, fixed on two beams, carries via three rods B a ring C. This ring supports by means of three flexible rods D a three-armed yoke E fitted with a universal joint F. The wire is suspended from F. By the bending of the flexible rods the yoke can move radially over a distance equal to the radial

play s between pin G and the periphery of a hole in the mounting plate. The flexible rods can be moved vertically and their length accurately adjusted; in this way the system as a whole can be given the required lateral stiffness and the pin made to lie exactly in the centre of the hole in the position of rest.

If we compare this arrangement with that of fig. 2*a*, we notice that the universal joint corresponds to the point of the wire which, in the Charron system, is at the height of the ring B. The hole in the mounting plate fulfils here the function of the ring, so that in the new construction the play s is equivalent to the play d in the old. The use of the universal joint avoids the friction and bending which the wire undergoes in Charron's construction at the level of the ring, thus greatly reducing the risk of wear or fatigue failure. The three flexible rods, like Charron's wire length AB, bear the weight G' of the pendulum bob etc. and, via the yoke, exert on the joint a centrally directed force which is proportional to the radial displacement. At a displacement u the above-mentioned point of Charron's wire is subjected to a force of magnitude Gu/l'. By appropriately dimensioning the flexible rods in the new construction, the same magnitude of restoring force can be produced.

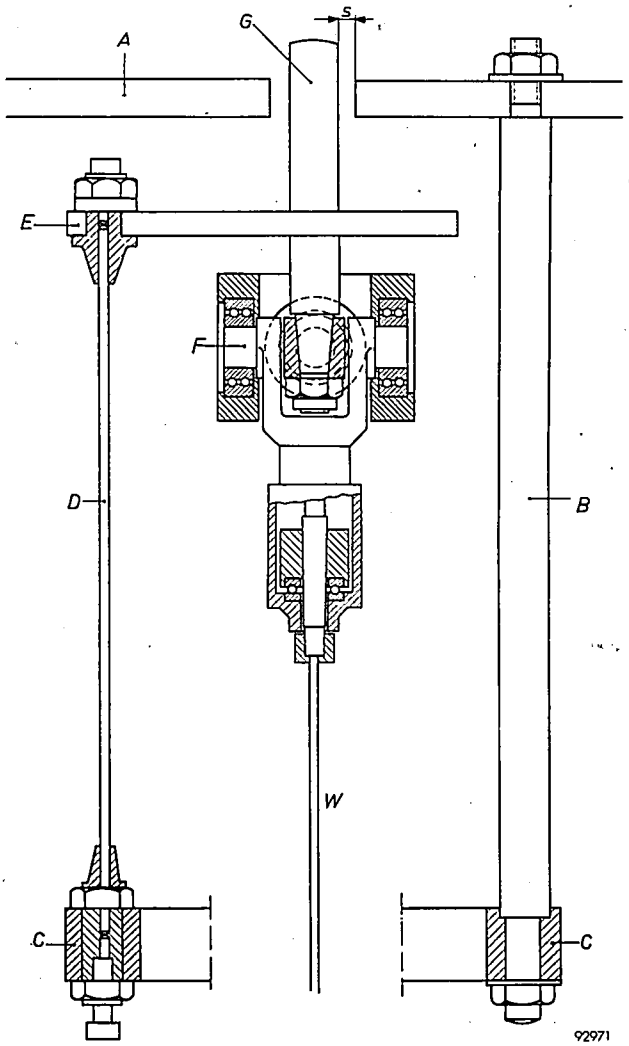


Fig. 4. Schematic diagram of the suspension system of the Foucault pendulum in the United Nations building. Mounting plate A carries three rigid rods B to which a ring C is fixed. Three flexible rods D, clamped on the latter, support a yoke E, which is fitted with a universal joint F. Pin G mounted on E, lies centrally in a hole in the mounting plate, and limits the lateral displacement of the yoke to the small distance s. The suspension wire W is fixed to the joint by clamping in a bifurcated steel cone. The dimensions of the flexible rods are such as to ensure the required lateral stiffness and to ensure that the bending stress in these rods is below their fatigue bending strength.

In connection with the dimensions of the flexible rods it should be borne in mind that the lateral stiffness c is determined not only by the dimensions and the modulus of elasticity but also to a large extent by the compressive force $\frac{1}{3}G'$. The length l' no longer corresponds to any physical quantity such as the length of the flexible rods, but is equal to $G'/3c$.

The lateral stiffness for this case (parallel-constrained ends) is given by the formula

$$c = \frac{P}{l_1} \frac{\frac{1}{2}ql_1}{\tan\frac{1}{2}ql_1 - \frac{1}{2}ql_1}, \quad \ldots\ldots \quad (4)$$

where $q^2 = P/EI$ and $P = \frac{1}{3}G'$. Further, l_1 is the length of the flexible rods, E their modulus of elasticity and I their second moment of area in bending.

The suspension wire is of hard drawn stainless steel with a diameter of 2.5 mm. The wire is clamped at both ends by a truncated cone of hardened steel with internal teeth. The cone has diametric saw cuts at its narrower end and is drawn fast into a ring by the weight of the bob.

In the design of the driving mechanism, which is necessary to keep the pendulum swinging without any reduction in amplitude, account had to be taken of the fact that access to the point of suspension is extremely difficult once the pendulum is mounted. The drive is therefore provided at the bob of the pendulum. The method adopted was to repel the bob by means of eddy currents generated in it by a coil mounted on the pedestal (*fig. 5*), the coil being energized at suitable moments by alternating current. If a copper plate is placed above such a coil at right angles to its axis, voltages will be induced in the plate which are 90° out of phase with the field, and currents which are in their turn almost 90° out of phase with these voltages. The eddy currents are therefore almost 180° out of phase with the current in the coil, thus giving rise to repulsion. If a not too large, horizontal, round plate is placed with its centre-point exactly in the axis of the vertical coil, the resulting force acting on the plate is a vertical one, but if the plate is slightly eccentric to the coil axis, the force will also have a horizontal component which can be used for driving the pendulum. A plate



Fig. 5. The pedestal, designed by architect G. Rietveld, on which the drive coil is mounted. The pedestal bears the inscription:

It is a privilege to live
This day and to-morrow.
Juliana

of this kind is contained in the lower half of the pendulum bob. In order to concentrate the magnetic field on the plate, 18 U-shaped yokes of ferroxcube are fitted around the coil.

This method, then, avoids the use of iron in the ball, and thus prevents the pendulum from being influenced by the magnetic field of the earth or by the steel structure of the building.

Some remarks may be added on the transfer of energy in this driving system. Eddy currents in the copper plate react on the coil with the result that the self-inductance of the latter is smaller than when the plate is absent. When the plate moves away this self-inductance increases and the energy transferred to the plate is approximately equal to $I^2 \Delta L$, where ΔL is the change in the self-inductance and I the current. This does not take into account that I itself changes slightly, but, partly because of the circuit arrangement (see below), this relation holds fairly accurately.

The moment at which the coil is energized is determined by the moment at which the rod in the yoke (*G*, fig. 4) looses contact with the mounting plate [4]). For this purpose, rod and hole were designed to constitute an electrical contact. A delay circuit, which also controls the duration of energization, ensures that the current in the coil is switched on at a specific time after the moment the rod leaves the mounting plate, viz. at approximately the same moment at which the centre of the bob passes the axis of the coil. The anode load of tube *II* in the circuit (see *fig. 6*) includes the D.C. winding of a transductor *T* [5]). The A.C. winding L_1 of this transductor is connected in series with the drive coil L_2 and a capacitor C''. In addition, another capacitor C' is connected in parallel with L_1. When the D.C. coil of the transductor is not energized, the circuit L_1-C' is in resonance and represents such a high impedance that the current through L_2 is only 85 mA.

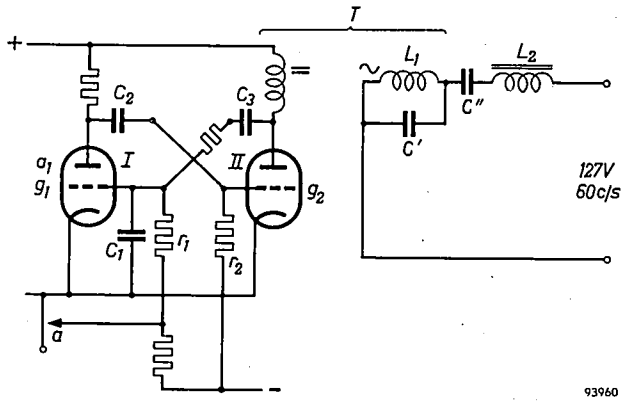


Fig. 6. Control and drive circuit for driving the pendulum. The control circuit (left) causes the required alternating drive current from the mains to be supplied to the drive coil L_2 only after a certain delay after the opening of contact *a*; the circuit sustains the current for a certain fixed time, even if contact *a* is again closed in the meantime. Contact *a* is formed by rod *G* (fig. 4) and the hole in the mounting plate *A*. In the quiescent state, tube *I* is conducting and tube *II* cut off. When *a* opens, the potential of g_1 falls and the potential of a_1 rises: at a certain point the circuit triggers, tube *II* now becoming conducting (via C_2) and tube *I* being cut off. The delay between the opening of *a* and the triggering of the circuit depends on the product C_1r_1. The anode current of tube *II* flows through the D.C. winding of transductor *T* and reduces the value of L_1 such that the drive circuit (right) comes into series resonance. After an interval, which depends on the products C_2r_2 and C_3r_1, the control circuit returns to the quiescent state and L_1 rises again. The value of C' is such that resonance now occurs in the parallel circuit L_1-C'. The latter then represents such a high impedance in the circuit that the current is reduced to an ineffective value.

[4]) R. Stuart Mackay, Amer. J. Phys. **21**, 180, 1953, describes a pendulum driven by eddy currents, the synchronization being derived from the pendulum bob.

[5]) A transductor (saturable reactor) is a choke with an iron core, the self-inductance of which can be varied by changing the magnetization of the iron core with the aid of a D.C. winding.

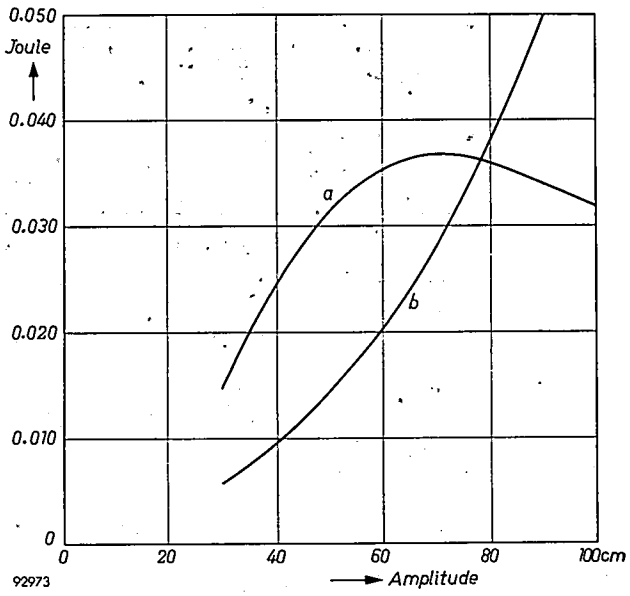


Fig. 7. The energy supplied per period to the pendulum (curve *a*) and that lost by air resistance (curve *b*) as a function of amplitude of swing. At an amplitude of approx. 80 cm the curves intersect. At smaller amplitudes the energy supplied is greater than that lost, i.e. the point of intersection represents a stable state of equilibrium.

When the transductor is energized by the anode current of tube *II*, the self-inductance of L_1 is reduced by about a half. The capacitance of C'' is such that the whole circuit now comes into series resonance, whereupon the current through L_2 rises to 240 mA. The ratio between operating and quiescent currents does not seem particularly large, but it must be remembered that the energy transferred is proportional to the square of the current. In this way about 0.035 joule is supplied to the bob in each period, which is sufficient to provide an amplitude of swing of the required value, viz. approximately 80 cm (see *fig.* 7).

Summary. A description is given of the suspension and driving systems for the Foucault pendulum suspended in the hall of the United Nations building in New York. The pendulum was presented on 7th December 1955 to the Chairman of the General Assembly by the Dutch Foreign Minister H. Luns on behalf of the people of the Netherlands. The main problem of preventing the pendulum bob from describing an ellipse instead of swinging in a flat plane was solved by utilizing the principle described by Charron. The construction is so designed as to reduce to a minimum the risk of the wire breaking owing to wear or fatigue. The drive, which is necessary to compensate for the energy losses, caused mainly by air resistance, is effected by means of a magnetic coil with ferroxcube core placed centrally under the pendulum. This coil is energized by alternating current and induces eddy currents in a copper plate contained inside the lower half of the bob. The energizing current is controlled by an electronic relay, which ensures that the current is switched on some time after the pendulum itself has broken an electrical contact in the suspension system, and keeps it switched on for a certain period. The coil is mounted on a pedestal designed by the architect G. Rietveld.

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.

2482: C. A. de Bock, J. Brug and J. N. Walop: Antiviral activity of glyoxals (Nature **179**, 706-707, April 6, 1957).

In screening compounds for antiviral activity against influenza virus, a number of α-keto-aldehydes appeared to be active. The test compounds were injected in the allantoic cavity of 11 day embryonated hen's eggs, followed after 1 hr by the virus. After incubation for 48 hours the hæmagglutination (H.A.) titre of the allantoic fluid was estimated. A compound was considered active if the difference between the logarithm of this H.A. titre and that of a control was >0.6. Active compounds and the difference in log H.A. titre were CH_3COCOH (2.0), p-OH—C_6H_4COCOH (2.0), p-OH.m-NO_2—C_6H_3COCOH (2.3), p-Br—C_6H_4 COCOH (0.7), m-NO_2—C_6H_4COCOH (2.1). The virus loses infective power when incubated with low concentrations (0.002 M) of glyoxals. At higher concentrations the glyoxals destroy the enzymatic activity of the virus. The virucidal action of the glyoxals is strong enough to explain the activity in the test.

2483*: H. de Lange Dzn.: Attenuation characteristics and phase-shift characteristics of the human fovea-cortex systems in relation to flicker-fusion phenomena (Thesis Delft, June 5, 1957).

The well-known frequency-response technique of systems analysis, in which the ratio of output amplitude to sinusoidal input amplitude is plotted against frequency in so-called attenuation characteristics, can successfully be applied to investigate the dynamic nature of the human visual organ from the retina up to the brain, using sinusoidally modulated light. This is possible because Talbot's law shows that, at flicker-fusion, the brightness-system works linearly. The internal threshold-value for flicker-fusion is supposed to be invariable with frequency at constant mean luminance and acts as a constant output value; the ripple-ratio $r =$ (amplitude of sinusoidal light variation)/(mean luminance) acts as a variable input of the brightness-system. This manner of investigation, previously applied in earlier papers with white light, is extended over nearly the whole range of cone-vision and is continued with coloured light. The existing theories on flicker-fusion provide no explanation for the shape of the attenuation characteristics obtained from the experiments and calculated from investigations by other authors. The attenuation charcteristics show at high luminance a pseudo-resonance effect, the bandwidth is greater and the slopes are such that for $r > 2\%$ the ratio (amplitude of fundamental)/(mean luminance) is decisive for flicker-fusion of any shape of interruption or modulation.

The well-known residual brightness-flicker just above the colour-flicker limit with heterochrome flicker-photometry with two anti-phase 100% sinusoidally modulated light beams can be reduced to zero by introducing an external phase correction $\Delta\psi$, which is found to be a function of luminance, colour difference and frequency. Subtracting logarithmically the attenuation characteristic of the colour-system from that of the brightness-system, it is found that the extra delay in colour-perception is identical to the filter-action of one integration process at high luminance; at low luminance a triple integration process occurs with the same time constant of 120 msec.

Using electrical analogues it is shown that the curves obtained are real attenuation characteristics of the brightness-system. In accordance with the pseudo-resonance peak in the attenuation characteristic for high luminance, an overshoot in the brightness perception occurs at about 1 c/s with perodic rectangular light impulses.

2484: A. Venema: The measurement of the pressure in the determination of pump speed (Vacuum 4, 272-283, 1954, No. 3, published Febr. 1957).

There is no commonly agreed method of determining the speed of a pump. One noteworthy method was proposed by Dayton in 1948 but it has not been accepted generally. Contributions to the problem made by other workers are briefly reviewed, followed by a close study of the basic elements of the definition of pressure. Special attention is given to the case where the number-density of the molecules varies along the mean free path. Such variation may occur at low pressures and exists, in particular, in the region of the pump mouth. The actual method of measuring pressure needs more consideration and it is shown that a gauge connected by a

tube to the system records the number of molecules per unit area arriving at the gauge-end of the tube. The conditions at the pump mouth have been investigated and an interpretation is given of the results of some measurements with regard to the distribution of the incident molecules. In conclusion, a method of measuring the speed of a pump is proposed which differs from Dayton's method. The new method is based on well-known concepts in the physics of gas flow at very low pressures. It has been put forward before, but simply as a postulate.

2485: M. H. de Lange: Heat transfer in glass furnaces from a theoretical and practical point of view (Travaux IVe Congrès int. du Verre, Paris, July 2-7, 1956, pp. 148-152; Chaix, Paris 1957).

An introductory survey is given of the various equations relevant to the calculation of the contribution of radiation to the transmission of heat through glass; certain equations relating to stationary conditions are also established. These equations are applied to the calculation of the vertical temperature distribution in a vat of glass. The equations are further applied to give an approximation to the effect of external cooling on the temperature of the inside wall of the furnace. An attempt is made to express the effect of radiated heat in non-stationary conditions, notably during the heating and the cooling of the glass.

2486: Y. Haven and J. M. Stevels: Note on the mechanism of ionic transport in glass (Travaux IVe Congrès int. du Verre, Paris, July 2-7, 1956, pp. 343-347; Chaix, Paris 1957).

Note drawing attention to information that can be derived from a comparison of diffusion and conductivity data, and to suggest that the mechanism of transport of Na^+ ions in glasses uses interstitialcies in certain cases (similar to the mechanism suggested by McCombie and Lidiard for Ag^+ in AgCl), whereas in other cases it is possible that vacancies are involved. The potential minima between the silicon-oxygen network available for the Na^+ ions are divided into sites — intersites and vacancies. Some examples are discussed which give some idea of the utility of the refinement of the theory given in this paper.

2487: A. Kats: The interaction of U.V. and X-rays radiation with silicate glasses and fused silica (Travaux IVe Congrès int. du Verre, Paris, July 2-7, 1956, pp. 400-411; Chaix, Paris 1957).

The formation of imperfections in silicate glasses and in fused silica under the action of X-rays has been studied. Optical absorption bands are observed in the region 2000-10 000 Å; they are attributable to centres which have captured electrons or holes. A number of alkali silicate glasses were irradiated at low temperature and measured at low temperature (80 °K). It is shown experimentally that the final result of the irradiation (either by u.v. of sufficient energy or by X-rays) is the loss of an electron by some of the oxygen ions or a displacement of the latter. It is still unknown via which intermediate transition states the centres are formed. Paramagnetic resonance measurements show the existence of electrons and holes attributable to the same centres as found optically. Imperfections in fused quartz after irradiation have also been investigated. It is shown that these may be due to centres originating from aluminium impurities.

2488: G. Diemer and W. Hoogenstraaten: Ambipolar and exciton diffusion in CdS crystals (Phys. Chem. Solids **2**, 119-130, 1957, No. 2).

The diffusion of photoconductivity in non-illuminated parts of CdS single crystals has been studied both on unactivated samples and on samples activated with Ag. Diffusion lengths ranging from a fraction of a micron to several hundreds of microns were observed. Measurements of the spectral response, the temperature dependence of the diffusion length and the P.E.M. (photo-electro-magnetic) voltage make it probable that the large values of the diffusion length are due to exciton diffusion. An activation energy of about 0.1 eV was found for thermal ionization of the excitons. At room temperature, in most of the Ag-activated crystals the thermal diffusion of the conductivity is a combination of ambipolar and exciton diffusion, a theory of which is given for a one-dimensional case.

2489: O. Reifenschweiler: Ionenquellen für kernphysikalische Untersuchungen (Elektrotechnik und Maschinenbau **74**, 96-103, 1957, No. 5). (Ion sources for nuclear research; in German.)

Survey of the more important ion sources used in nuclear research. The various types of sources are described, against the background of their common working princple. Some results of work on H.F. ion sources delivering currents of the order of 10 mA are given.

2490: N. W. H. Addink: Excitation energies in line spectra (Spectrochimica Acta **9**, 158-159, 1957, No. 2).

The calculation (according to Boltzmann's law) of the number of excited ions may not be carried out on the basis of simply adding the energies of ionization and excitation. The percentage of ions must first be determined (following Saha) and Boltzmann's distribution law then applied to the number of ions so found.

2491: S. Woldring: Continue onbloedige bloeddrukmeting bij de mens (Ned. T. Geneesk. **101**, 949-952, 1957, No. 20). (Continuous non-bloody blood-pressure measurement in man; in Dutch.)

Blood pressure in the arteries of the hand is measured on the principle of the "relaxed arterial wall". The pressure gradient established by the elastic tension of the arterial wall is overcome by compression of the surrounding tissues in a plethysmograph-like system, connected to a low-compliance manometer. Sample records of blood pressure under varying circumstances are given.

2492: J. H. N. van Vucht: Beitrage zur Kenntnis des Systems Cer-Aluminium (Z. Metallk. **48**, 253-258, 1957, No. 5). (Contribution to the study of the system cerium-aluminium; in German.)

The system Ce-Al was investigated by metallographic, thermoanalytical and X-ray diffraction methods for cerium concentrations above 50 at. %. Neither the compound Ce_2Al, nor the compound Ce_3Al_2, could be affirmed. Instead we discovered the existence of two modifications of a compound Ce_3Al. Below 230 °C, Ce_3Al has a hexagonal Ni_3Sn structure, above that temperature it is cubic with Cu_3Au structure. The compound CeAl was indexed as an orthorhombic one. A table with observed spacings and intensities is given.

2493: H. T. Schaap: De waterstofziekte van koper (Metalen **12**, 204-208, 1957, No. 11). (Hydrogen embrittlement of copper; in Dutch.)

A survey, with some results of the author's experiments, is given of the well-known effects of hydrogen embrittlement of copper. Copper containing more than 0.01 weight % oxygen becomes brittle when heated in hydrogen above 400 °C. Hydrogen diffusing into copper containing inclusions of Cu_2O reduces this oxide to metallic copper and water vapour. High pressures of water vapour are built up in this reduction causing local rupture along grain boundaries and resulting in embrittlement of the metal as a consequence of the fact that the rate of diffusion of hydrogen in copper is much higher than that of water vapour. In addition to these phenomena hydrogen embrittlement shows several other aspects, e.g. (*a*) cracks on the surface, (*b*) change of dimensions, (*c*) blisters on the surface, (*d*) series of holes along the grain boundaries, (*e*) holes in the interior of the grains. The conditions under which hydrogen embrittlement is encountered are described. Oxygen must be present as Cu_2O or as oxides from foreign metals whose affinity for oxygen is not much greater than that of copper and which show a measurable solubility in that metal. Our own experiments make it doubtful whether oxygen in solid solution may cause embrittlement. Small quantities of water vapour in the gas atmosphere can only be harmful in the presence of a material dissociating the water vapour, for instance chromium.

Now available

F. M. Penning: Electrical discharges in gases, Philips Technical Library, pp. viii+75, 29 figures.

This book is a translation of the Dutch original, which appeared in 1955, two years after the author's death. Although in recent years several new books have appeared on the subject of gas discharges, it is felt that the present book nevertheless fulfils a need, offering as it does a concise synopsis which can profitably be used by students as a basis for further study.

The contents are as follows: 1. Gas discharges, natural and man-made; 2. The conduction of electricity in metals and gases; 3. Non-self-sustaining discharges; 4. The movement of electrons and ions through a gas; 5. The non-self-sustaining arc discharge; 6. The Townsend discharge and breakdown; 7. Sparks and lightning; 8. The glow discharge; 9. The self-sustaining arc discharge; 10. The positive column. A bibliography and an index complete the book.

VOLUME 19, 1957/58, No. 9 pp. 245-272 Published 27th February 1958

Philips Technical Review

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

MONITORING, CONTROL AND SAFETY EQUIPMENT FOR A NUCLEAR REACTOR OF THE SWIMMING-POOL TYPE

I. GENERAL DESCRIPTION

by M. van TOL.

621.316.7:621.039.4

On 28 June 1957 a nuclear reactor was put into operation for the first time in the Netherlands, at the exhibition "Het Atoom" at Amsterdam. While the reactor is of American manufacture (A.M.F. ATOMICS, New York), the Netherlands industry played a considerable part in its planning and installation. Most of the electronic control and safety equipment, for example, was supplied by Philips.

Part I of an article on this equipment is printed below and contains a brief explanation of the functioning of a nuclear reactor and a general description of the apparatus. Part II, to be published in a later issue, will deal in detail with certain components, including one of the neutron flux meters, the gamma-ray indicator and the safety amplifier.

The "nucleus" of the exhibition "Het Atoom" [1], which was held in Amsterdam during the summer months of 1957, was undoubtedly the experimental nuclear reactor that was shown in operation (*fig. 1*). The reactor, which was designed and supplied by A.M.F. ATOMICS Inc., New York, a subsidiary of the American Machine and Foundry Inc., is soon to be transferred to the Technische Hogeschool at Delft.

The principal data concerning this reactor are given below (reproduced by courtesy of A.M.F. ATOMICS). Some of the terms occurring here and also further in the article are explained under the heading "Operation of a nuclear reactor" (see pp. 246-248).

Type:
"Swimming Pool" type, i.e. the reactor core is set up in a water-filled tank, the water having the three functions mentioned below. The tank is in the form of a well (7 m deep, 3 m diameter; see *fig. 2*).

Power:
max. 100 kW (at the Exhibition the reactor was operated at max. 10 kW).

Fuel:
enriched uranium (20% U^{235}).

Moderator:
ordinary (not heavy) distilled water.

Coolant:
the same water as above.

Shielding:
the same water as above, and concrete.

Dimensions of reactor core:
$38 \times 45 \times 60$ cm.

Neutron flux inside reactor core:
max. 10^{12} neutrons per cm^2 and per second.

Coarse control and safety elements:
four shim and safety rods of boron carbide (three at the Exhibition).

Fine control:
one control rod of stainless steel.

Excess reactivity:
0.5% at the Exhibition (will be more later).

Max. control effect of all rods together:
12.5%.

The latter two items indicate that the factor k by which the number of neutrons rises or falls from one generation to another amounted at maximum

[1] Atomics and Nuclear Energy 8, 289, 1957 (No. 8); Nuclear Power 2, 334, 1957 (No. 16).

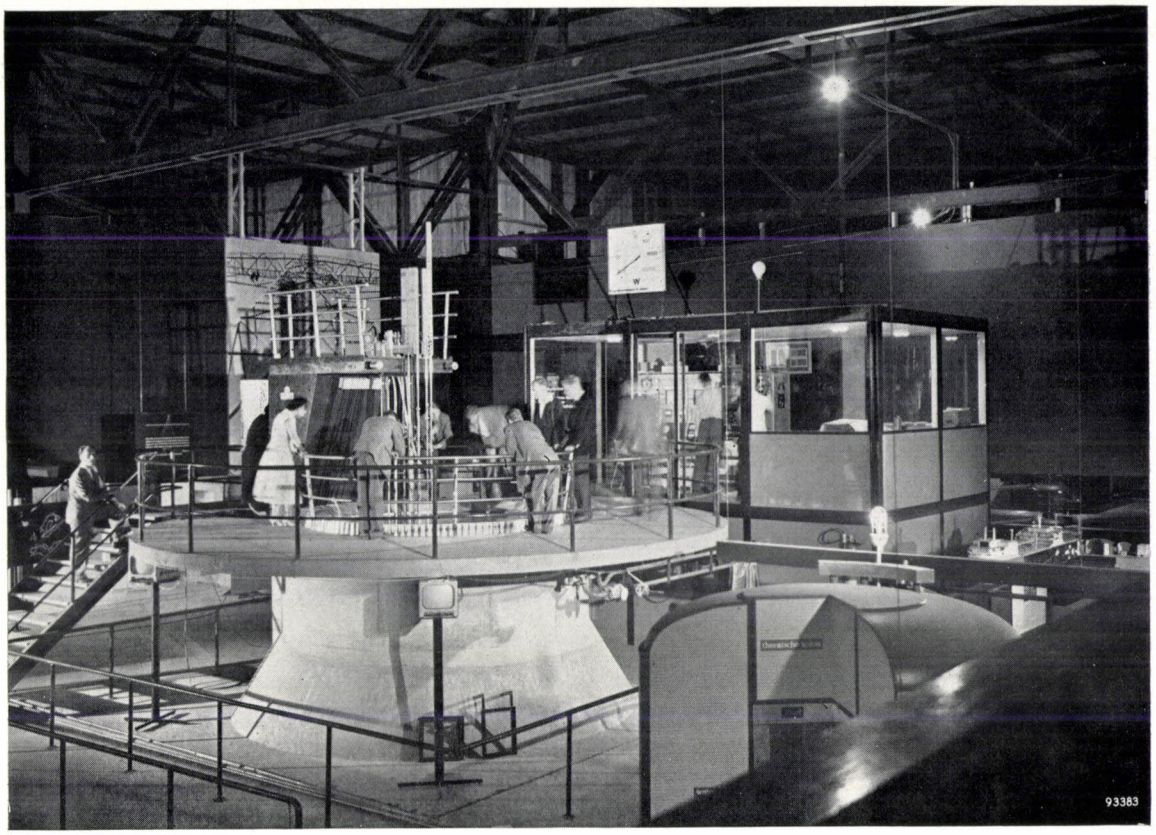


Fig. 1. Experimental nuclear reactor of the swimming-pool type at the exhibition "Het Atoom" in Amsterdam (Schiphol Airport). The photograph shows the upper part of the cylindrical tank, around which a platform has been fitted for visitors. The monitoring, control and safety equipment is installed in the cabin in the background.

to 1.005 (all rods up) and at minimum to 12.5% less, i.e. about 0.9 (all rods down).

The reactor core is near the bottom of the tank (fig. 2). The fuel elements, which contain altogether 18 kg of uranium, are entirely sheathed in aluminium to prevent the entry of highly radioactive fission products into the water. The water can circulate freely between the fuel elements. Some of the elements are designed to allow a safety or control rod to be moved up and down inside them. This is done with the aid of servomotors mounted on a platform above the tank. The water is kept circulating by a pump. It is passed through a mineral extractor (ion exchanger) to remove as many as possible of the mineral constituents which, in the long run, would "poison" the moderator.

Operation of a nuclear reactor

We shall first briefly discuss the operation of a nuclear reactor which uses uranium as "fuel"; in doing so we shall clarify some of the terms mentioned in the course of this article.

The operation of a uranium reactor is based principally on the properties of the uranium isotope U^{235}. Natural uranium contains only about 0.7% U^{235} and more than 99% U^{238}. Use is often made of "enriched" uranium, the U^{235} content of which is, for example, 20%.

Fission of U^{235} nuclei

After capturing a neutron, a U^{235} nucleus can split into two (rarely, three) fragments. The fragments are mostly unequal in size, the one having a mass number of about 95 and the other about 140. With each fission an energy of some 200 MeV is released, largely in the form of the kinetic energy of the fission fragments. By the intermediary of a coolant, for example, this energy can be extracted from the reactor and put to use. A small proportion of the fission energy is released as gamma radiation; a shield, consisting perhaps of water and concrete, ensures adequate absorption of this radiation.

The fragments of a single fission produce, on an average, two or three neutrons, and these in their turn can be used to split other U^{235} nuclei and so to keep the process self-sustaining. However, the neutrons liberated from the fission products have very high velocities (energy in the order of 1 MeV), and in uranium having 20% U^{235} such fast neutrons are not able to split U^{235} nuclei in sufficient numbers. To do so they must be slowed down to "thermal" velocities corresponding to about 0.03 eV. (This is expressed by saying that the U^{235} nucleus has a small effective cross-section for fast neutrons and a large one for "thermal" neutrons.) The slowing-down of the fast neutrons in the reactor, without excessive neutron absorption, is performed by a "moderator". This consists of a substance (e.g. graphite, ordinary or heavy water) which is mixed with the uranium, or which surrounds it, in such a way that the fast neutrons enter it and lose velocity by collisions. After only 18 collisions with H atoms, for instance, a neutron with an initial energy of 1 MeV has become a thermal neutron.

Chain reaction

A number of the neutrons liberated by nuclear fission are lost; some escape from the reactor, others are absorbed in U^{238} nuclei, in the moderator, in the fission products or in deliberately introduced neutron-absorbent material (see below). It has proved readily possible, however, to create conditions in which a sufficient number of the liberated fast neutrons can be slowed down to thermal velocities in order to sustain the reaction (chain reaction). Once these conditions are realized, the reaction can set in spontaneously, for there is usually a thermal neutron already present (e.g. from cosmic radiation) which can serve as the "trigger". To have a measurable quantity from the start, however, an artificial source of neutrons is always used in practice for supplying the initial neutrons.

The larger the reactor, the smaller is its surface area in relation to its volume and thus the smaller in proportion is the chance of neutrons escaping. There is, in fact, a *critical size*, below which no chain reaction can take place in the reactor because of excessive neutron losses. The magnitude of this critical size is strongly dependent upon a variety of circumstances, such as the configuration, the nature of the moderator, the U^{235} content and the temperature.

Neutron conservation

In a working reactor neutrons are constantly being liberated and constantly absorbed or captured. We can therefore speak of neutron generations having a certain average lifetime. An important quantity is the *multiplication factor* k, which is the ratio of the number of neutrons in a given generation to the number in the preceding generation. The condition for a just self-sustaining chain reaction (stationary state) is that k be equal to unity. To start a reactor, k must be greater than unity; a very small excess, for instance 0.1%, i.e. $k = 1.001$, may be sufficient.

Although two to three new neutrons are produced with each fission, so that if there were no losses the factor k would be 2 or 3, the maximum value that k can reach in an actual reactor is usually not more than 1.01. The reasons for this are the following:

1) Escape of neutrons from the reactor.
2) Absorption of neutrons in the moderator, in the fission products and in U^{238}, in particular resonance absorption during the slowing-down process in the moderator: the chance that neutrons having an energy of 6.5 eV are captured by U^{238} nuclei ("resonance") is not to be neglected.

The loss of neutrons is slightly offset by the fact that a few of the neutrons not yet slowed down also cause the fission of U^{235} and U^{238} nuclei, giving rise to more than one fast neutron per fission. This is the so-called fast fission effect.

When a state is reached where k is greater than unity — which is more easily reached the richer the uranium is in U^{235} — the neutrons then continuously increase in number, and so therefore do the number of fissions per second and the power

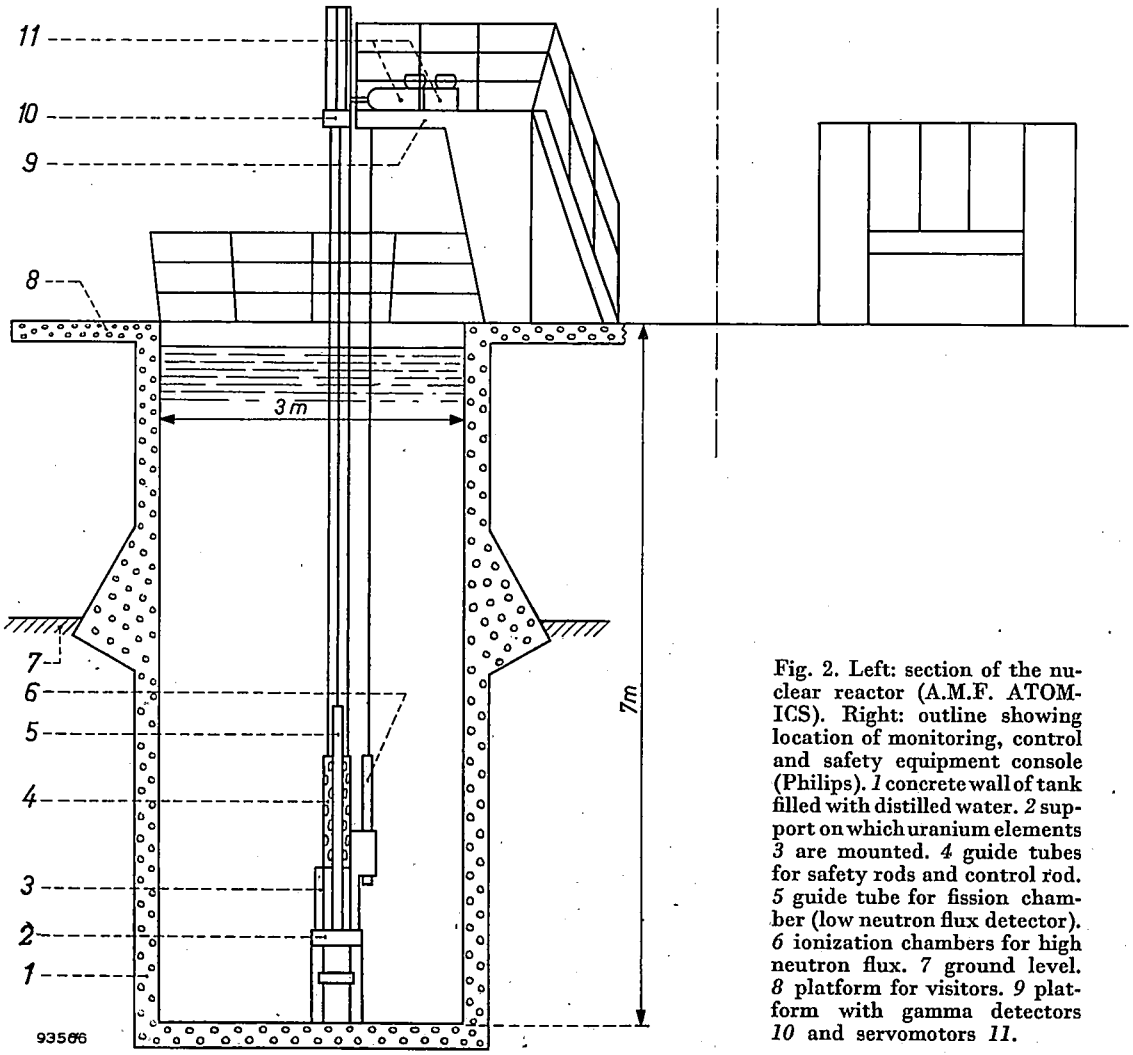


Fig. 2. Left: section of the nuclear reactor (A.M.F. ATOMICS). Right: outline showing location of monitoring, control and safety equipment console (Philips). *1* concrete wall of tank filled with distilled water. *2* support on which uranium elements *3* are mounted. *4* guide tubes for safety rods and control rod. *5* guide tube for fission chamber (low neutron flux detector). *6* ionization chambers for high neutron flux. *7* ground level. *8* platform for visitors. *9* platform with gamma detectors *10* and servomotors *11*.

delivered by the reactor. The time required for the power to increase by a factor e — the time constant — is known as the "period" of the reactor.

Control

It is essential to be able to control the rate of the reaction in a nuclear reactor. Since the fission itself and the subsequent slowing-down in the moderator take place in such an extremely short time, it might be thought that each chain reaction started would spread too rapidly to allow timely intervention by any regulating mechanism and hence would make it impossible to control the reaction. The fact that control is possible is due to the circumstance that not all the neutrons are immediately liberated during the fission. Of the total neutrons liberated, 99% are indeed "prompt" in this sense, but the gradual release of the remainder can still be observed for minutes after a fission process. These "delayed" neutrons make it possible to control the reaction by mechanical, that is to say relatively slow, means.

As we have seen, with each fission an energy of about 200 MeV is liberated. The power delivered by a reactor is proportional, then, to the number of fissions per second. This number in its turn is proportional to the number of thermal neutrons present. In order to measure (and also to control) the power of the reactor it therefore suffices to measure the number of thermal neutrons. This is possible because the neutrons are in continuous motion through the reactor. The product of the number of thermal neutrons per unit volume and their average velocity (this velocity is not dependent on the power developed by the reactor) is called the *neutron flux* (N). This product, which is proportional to the reactor power, is one of the main quantities constantly monitored in a reactor and is employed for control purposes.

If u is the number of U^{235} nuclei per unit volume, σ the effective cross-section of a U^{235} nucleus for thermal neutrons, and V the volume of the reactor, then $Nu\sigma V$ is the number of fissions occurring per unit time. With each fission 200 MeV is released, i.e. 3.2×10^{-11} joule. The power developed, P, is therefore:

$$P = Nu\sigma V\times 3.2\times10^{-11} \text{ watt.}$$

With a reactor in the stationary state the multiplication factor k is, by definition, precisely equal to 1. Naturally, this state can never be maintained exactly; the number of neutrons (hence the neutron flux and the power) gradually increases or decreases. In order nevertheless to keep the power within definite bounds it is necessary to have some means of varying k. For this purpose use is made of one or more rods of a material that absorbs neutrons (e.g. cadmium, boron, iron). By varying the depth to which these *control rods* are inserted into the reactor (by means of servomotors, for example) the neutron losses and thus the k factor can be varied. Some of the rods ("shim" rods) are usually employed for making coarse adjustments, being so positioned that k is equal to unity when the others, the control rods proper, are in their middle position; this makes fine control possible over a maximum range on either side. The shim rods are also sometimes used as "safety rods" which, in case of emergency, are quickly plunged into the reactor to bring it to a standstill. Control operations are usually carried out automatically, utilizing automatic control techniques.

A controlled chain reaction, with graphite as moderator and control rods of cadmium, was first achieved under the direction of Fermi in the University of Chicago in December 1942.

Some details of the fuel consumption and reaction products of a nuclear reactor are given in the appendix to this article.

Survey of the instruments required

To ensure the efficient functioning of the reactor certain quantities must be continuously measured (and some recorded) and it must be possible to operate the servomotors manually as well as by automatic control. The quantities measured are:

a) the neutron flux (at five points in the reactor core);
b) the temperature of the water (at three points);
c) the position of the control rod and of the safety rods;
d) the gamma radiation (at two points above the water and at one point near the ion exchanger).

These quantities are converted by various types of detectors into electrical signals which can be read directly from meters. The generated power, which is proportional to the neutron flux, and the "period" (time constant), which is derived from the rate of change of the neutron flux, can also be read from meters. A too short period indicates that the neutron flux (and hence the power) is increasing too rapidly.

The reactor power is regulated by moving the control rod up or down by means of a servomotor.

Of extreme importance are the provisions made for safety. There is no question of an atomic explosion; what must be guarded against are rather such dangers as excessive gamma radiation in the surroundings and damage to the reactor core by too high local temperatures. If such a danger threatens, an alarm must be given at once and the reactor put out of operation as quickly as possible ("scram"). This is done in the following way. The safety rods are suspended from their driving mechanism by means of electromagnets. If the reactor period is too short, the neutron flux too high, the gamma radiation excessive, or if certain other undesirable conditions arise, the current through the magnets is immediately cut off. The rods then drop to their lowest position, where they absorb so many neutrons that the reactor stops.

Nearly all the measuring, control and safety equipment for this reactor was supplied by Philips to the specifications of A.M.F. ATOMICS. The equipment, which is largely electronic, is mounted on five standard racks and a control desk, arranged as a console (*fig. 3*). The three middle racks contain the most important instruments, including the recording potentiometers [2]) for the neutron flux. The two photos in *fig. 4* show the five racks from the rear; amongst the things visible in these views are the mains voltage stabilizer and the complete relay

[2]) H. J. Roosdorp, An automatic recording potentiometer for industrial use, Philips tech. Rev. **15**, 189-198, 1953/54.

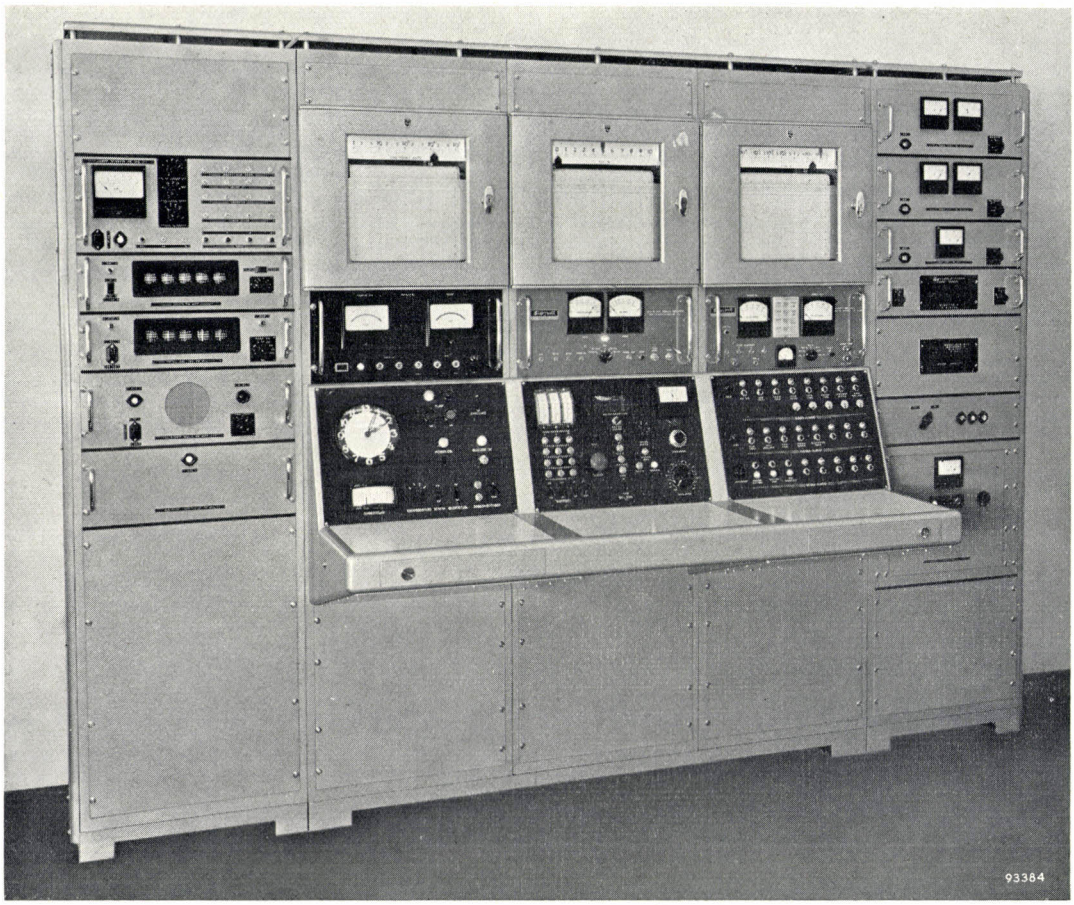


Fig. 3. Console carrying the monitoring control, and safety equipment. Left rack, from above: gamma indicator, pulse counter, pre-set time unit, loudspeaker unit, unit for temperature measurements. The rack on the right contains, among other things, the stabilized voltage sources. For the central section (three racks), see fig. 7.

system for operating the servomotors and the safety mechanism. A closer view of the relay system is given in *fig. 5*.

Neutron flux measuring channels

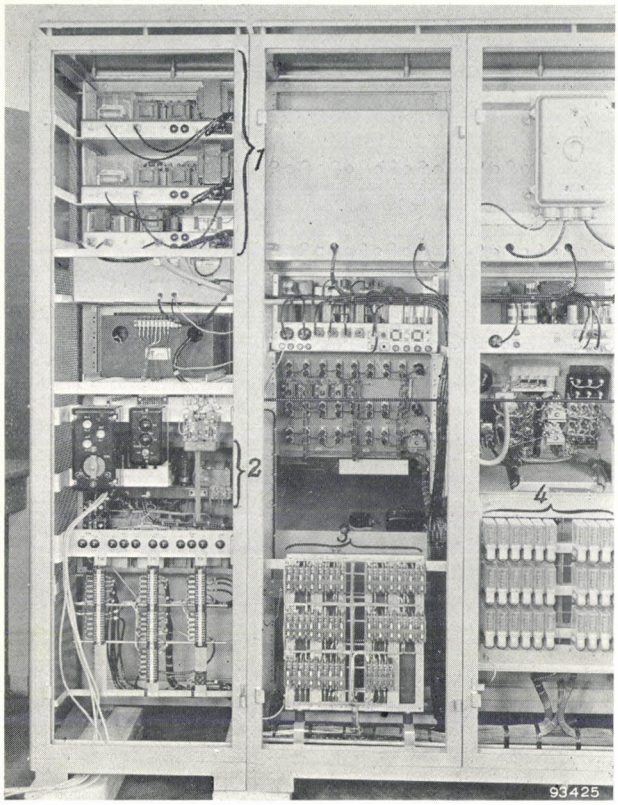
For measuring and automatically controlling the neutron flux, and also for safeguarding against the consequences of too high flux, it would in principle be enough to use one and the same detector, connected with an amplifier whose output voltage is a measure of the neutron flux. In actual practice, however, no less than five detectors are used, each with its own measuring channel. The reasons for this appear from the following considerations.

a) For *recording* purposes, a *logarithmic* scale is required, because the neutron flux from the initial state to that of maximum power increases by a factor of 10^8 to 10^9. Logarithmic flux measurement is also needed for determining the reactor period. This calls for complicated electronic equipment, capable of meeting high demands as to sensitivity and stability.

b) For *control* purposes, a *linear* measuring instrument is to be preferred. Provided the instrument has sufficient measuring ranges, it can also serve for accurately measuring the neutron flux (the accuracy, of course, would be far less with a meter having a single range covering many decades).

c) The *safety devices* must react not only when the neutron flux is too high, the period too short, the temperature excessive, etc., but also in the event of any fault whatsoever in the safety system itself. Reliability demands therefore a system as simple and robust as possible, and for this reason it cannot well be combined with the other channels. To increase the margin of safety, the safety channel is duplicated.

d) At very *low neutron fluxes* (less than 10^{-5} of the maximum flux) a detector of the kind used for normal flux would deliver too weak a current to be accurately measured. Nevertheless, accurate measurement is still necessary, if only because even at low fluxes the reactor period



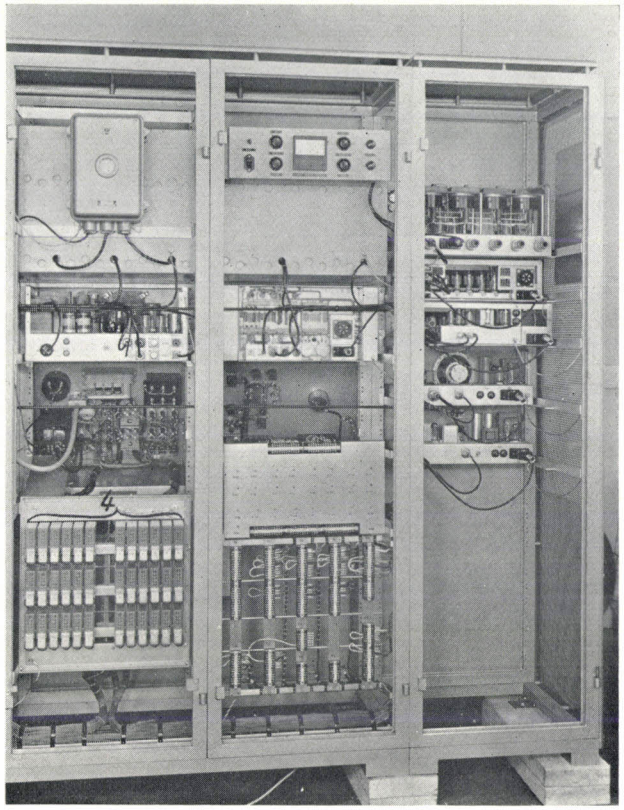


Fig. 4. Rear view of console. *1* three stabilized D.C. power supplies. *2* A.C. voltage stabilizer. *3* and *4* relay banks for servomotors and safety channels (see fig. 5).

must not be too short and must be constantly monitored. Hence the inclusion of a separate channel with different types of detector and amplifier for the measurement of low fluxes.

These considerations led to the following five channels for neutron flux measurement:

1) Channel for low flux, with logarithmic meter, recording apparatus and period meter.
2) Channel for high flux, with logarithmic meter, recording apparatus and period meter.
3) Channel for high flux, with linear meter and recording apparatus; also used for automatic control.
4) and 5) Two identical safety channels, designed with the emphasis on reliability, and having "fail-safe" features.

We shall now discuss the channels in the above order, after which we shall deal with temperature measurement, the position indication of the control and safety rods, gamma-ray measurement and the automatic control of the reactor.

Measuring channel for low neutron flux

The detector for low neutron flux is a so-called fission chamber (of American manufacture), which is a type of counter tube whose inside wall is lined with a fissionable material, in this case uranium. Fission of a uranium nucleus by a neutron results in two fast oppositely-moving fragments, one of which is almost certain to pass through the gas of the counter tube, producing ionization. (Of the

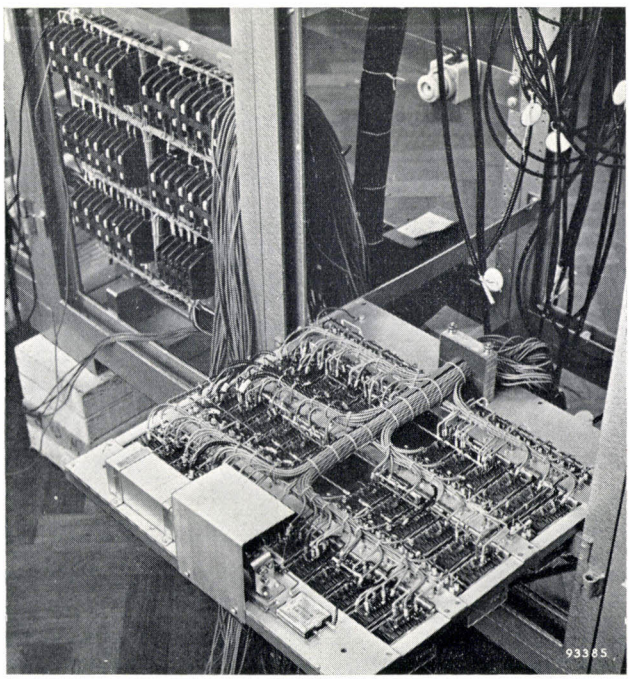


Fig. 5. Relay banks for servomotors and safety channels, one of the relay panels having been lowered for inspection.

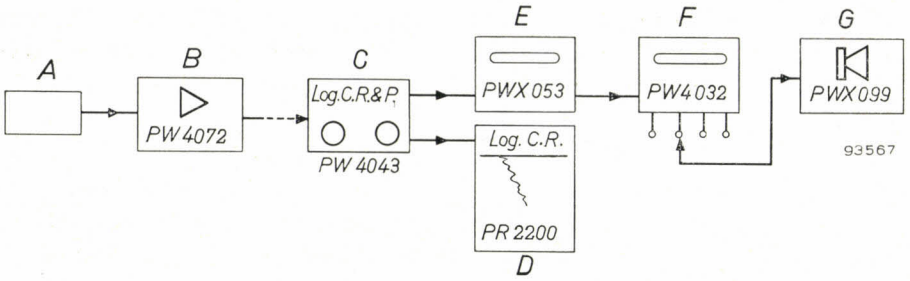


Fig. 6. Block diagram of measuring channel for low neutron flux. *A* fission chamber, *B* pulse amplifier, *C* logarithmic count-rate and period meter, *D* automatic recording potentiometer, *E* pre-set time unit, *F* pulse counter, *G* loudspeaker unit.

92 electrons surrounding the nucleus of a uranium atom, about 40 are stripped off after nuclear fission; thus the fission fragments are ions, each carrying some 20 positive elementary charges.) Consequently, almost every nuclear fission at the wall of the counter tube produces a current pulse.

The detector (*A* in *fig. 6*) is mounted in or near the reactor core and is connected to a linear pulse amplifier *B*, in which the current pulses are converted into voltage pulses and amplified 30 000 times. The amplifier is set up as close as practicable to the detector in order to minimize the capacitance of the connecting cable.

The output pulses, having an amplitude of a few volts, are fed to the logarithmic count-rate and period meter *C* ("log. C.R. & P."). Here they are separated by a discriminator ("Schmitt trigger") from the much smaller background pulses (caused by alpha particles produced in the counter by the radioactive disintegration of certain fission products) and then shaped into pulses of standardized amplitude and width. From these standard pulses a direct voltage V_1 is derived which is proportional to the logarithm of the number of pulses per second, and hence to log. neutron flux N. (Methods of obtaining this logarithmic relation will be discussed

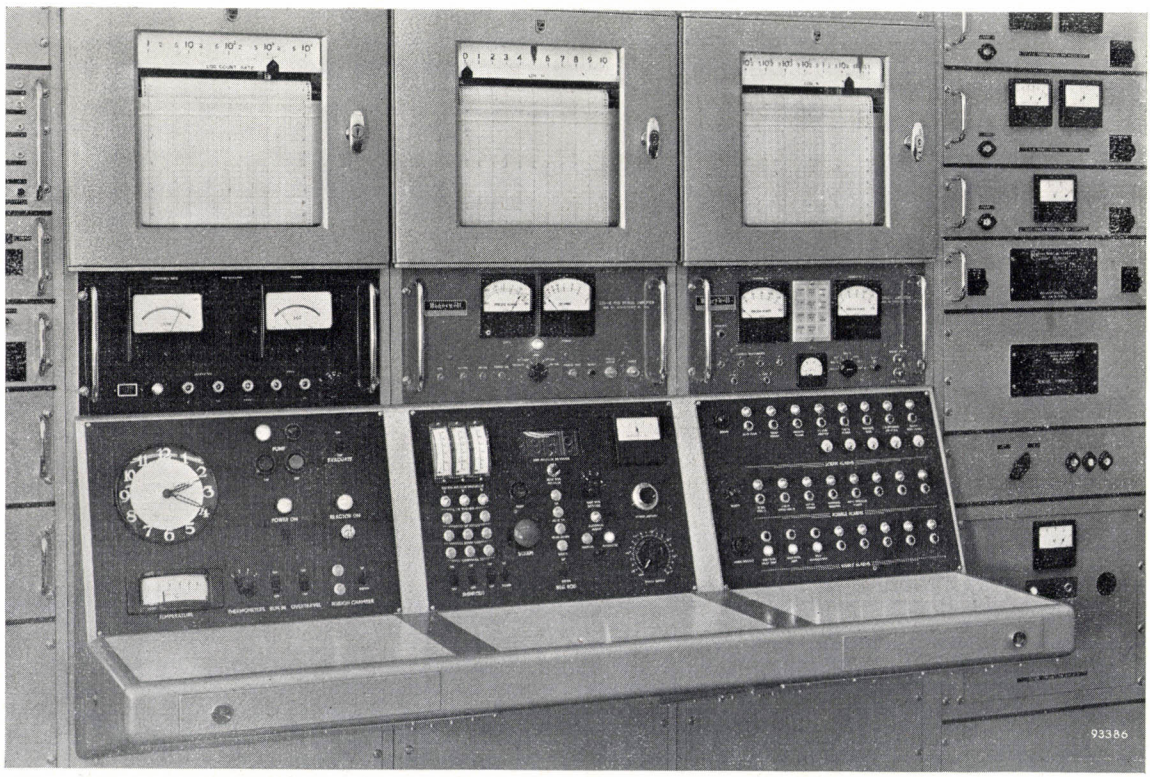


Fig. 7. Central section of console (cf. fig. 3). The three recorders above (from left to right) record respectively low neutron flux on a logarithmic scale ("log. C.R."), high neutron flux on a linear scale ("lin. N") and high neutron flux on a logarithmic scale ("log. N"). Below the recorders are mounted the associated flux, period and power meters.

Control panels: temperature meter below clock; the middle panel contains the position indicators for the safety and control rods and the "Power Demand" knob, and the right panel the pilot lamps.

in Part II.) The voltage V_1 is indicated on a voltmeter as well as recorded by an automatic potentiometer (*fig. 7*), both having a logarithmic scale from 1 to 10^4 pulses per second. From V_1 a differentiating amplifier derives a voltage V_2 which is proportional to

$$\frac{1}{N}\frac{dN}{dt},$$

that is to the relative change in the neutron flux per second. V_2 is thus inversely proportional to the period T (the time in which the flux increases by a factor e) and it is made to deflect the needle of a meter on which T can be read directly in seconds. The unit containing the logarithmic counting-rate and period meter is shown separately in *fig. 8*.

The actual counting is performed by five decimal counter tubes, type E1T [3]), mounted in unit PW 4032 (fig. 3). Fluorescent spots on the walls of these tubes together indicate a five-digit number which corresponds to the number of pulses counted in a given time. Apart from the direct indication in decimal digits, an advantage of this apparatus is its very rapid counting rate (resolution time 0.5 μsec) [4]).

The counting period can be pre-set with unit PW X053 (*fig. 9*). This contains two gate circuits, which can be simultaneously opened by means of a push-button. The one gate then passes the pulses to be counted to the counter circuit PW 4032; via the other gate, auxiliary pulses from the mains voltage (50 per second) are passed to a similar counting device in unit PW X053. As soon as a pre-set number of auxiliary pulses is reached, the second counting system closes both gates. The number of fission pulses counted in the time selected can now be read on the counter tubes of the first gate. The counting time, determined by the number of auxiliary pulses, can be adjusted in steps from 5 to 2000 sec (knob on the right in fig. 9). In this way very small values of neutron flux can still be accurately measured.

Finally, there is the loudspeaker unit PW X099 (fig. 3 and fig. 6). With this it is possible to make audible either all pulses counted by unit PW 4032 or only each tenth, hundredth or thousandth pulse. During start-up of the reactor, when the neutron flux is still low, the operator must watch several meters at the same time, for which reason an audible signal has proved very useful. A change in the average pulse rate can be heard immediately; a too rapidly increasing pulse rate means too short a period, indicating the necessity for prompt action.

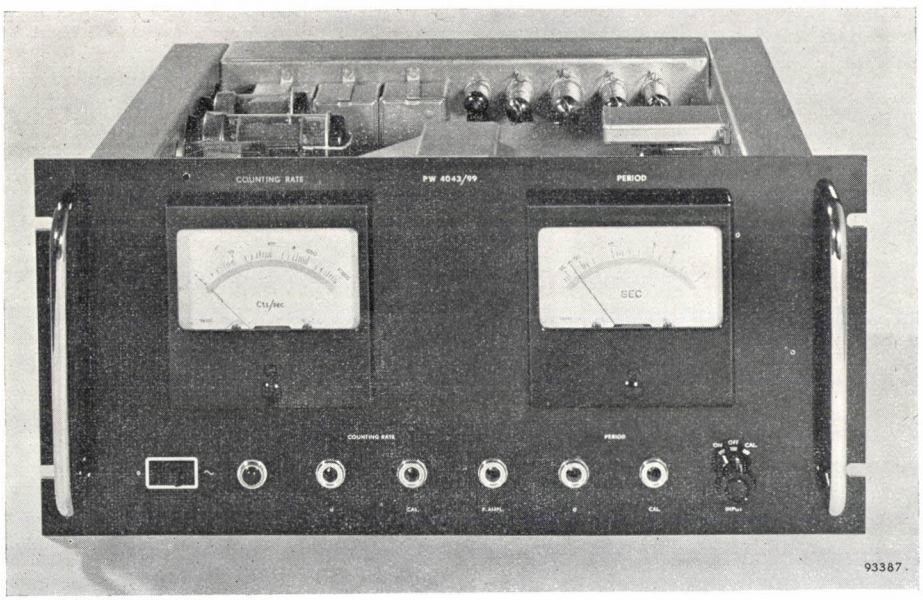


Fig. 8. Logarithmic count-rate and period meter PW 4043 for low neutron flux. The meter on the left indicates the number of pulses per second (scale 1 to 10^4), the meter on the right indicates the reactor period in seconds (scale from −30 via ∞ to 3 seconds; cf. fig. 10).

The operating personnel find the audible signal so valuable that they also make use of the low-flux channel even at high neutron-flux levels (the low-flux channel being the only one which provides an audible signal). For this purpose the fission chamber is raised sufficiently above the reactor to reduce the pulse rate to a usable value.

[3]) A. J. W. M. van Overbeek, J. L. H. Jonker and K. Rodenhuis, Philips tech. Rev. **14**, 313, 1952/53.

[4]) E. J. van Barneveld, Philips tech. Rev. **16**, 360, 1954/55.

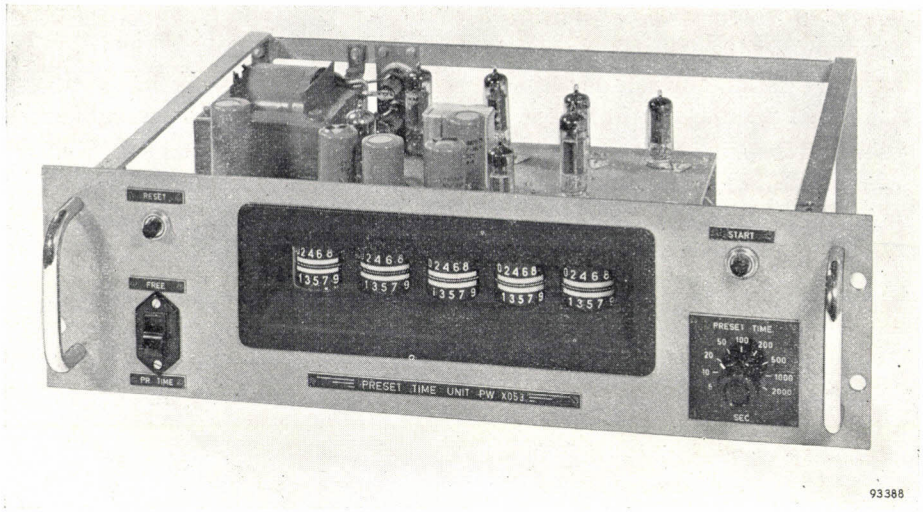


Fig. 9. Pre-set time unit PW X053 with five decade counter tubes type E1T. When the start button (top right) is pressed, this circuit begins to count auxiliary timing pulses (50 per second) and at the same time unit PW 4032 begins to count the pulses from the fission-chamber counter. As soon as the pre-set number of timing pulses has been reached (corresponding to a time of 5 to 2000 seconds, adjusted by the switch, bottom right), both counters stop automatically and the number of fission-chamber pulses counted can be read on unit PW 4032. If the switch, bottom left, is put in the "free" position, the PW 4032 unit goes on counting. The push-button, top left, serves for setting the counter PW X053 to zero.

Logarithmic measuring channel for high neutron flux

In the case of high flux levels, separate pulses are not counted as in the measurement of low neutron flux, but the average direct current of an ionization chamber is measured. To be more exact, what is measured is the difference current of the two parts of a differential (or "compensated") ionization chamber (of American manufacture). Only one of these parts is sensitive to neutrons, its wall being lined with boron carbide, but both are sensitive to the same degree to gamma radiation, which can be particularly intense in the reactor core. By taking the difference current, the effect of the gamma radiation is eliminated; the measured current is thus proportional to the neutron flux.

The difference current (50 μA at the nominal power) is fed to a logarithmic neutron flux meter (annex period meter) PW X081. As will be explained in detail in Part II, the grid-cathode space of an electron tube in a special circuit functions in this meter as a logarithmic element. In this way a voltage is obtained which is almost exactly proportional to the logarithm of the difference current over a very wide range (6 to 7 decades). After amplification a voltage is produced which is indicated on a meter and also recorded ("log. N recorder"). The scale, calibrated in %, runs from 10^{-4} % to 300 %, corresponding to difference currents from 5×10^{-11} A to 1.5×10^{-4} A. The first decade (10^{-4} to 10^{-3} %) of this instrument coincides with the last decade (10^{3} to 10^{4} pulses per second) of the meter for low neutron flux.

As in the case of the channel for low flux, a differentiating amplifier delivers a voltage which is inversely proportional to the period T; the latter can be read directly from a meter in seconds (*fig. 10*).

The supply voltages needed for the ionization chambers are a direct voltage of 600 V positive with respect to earth, and one of max. 350 V negative with respect to earth, the latter being adjustable in steps of approximately 35 V. Both voltages are stabilized and are taken from the power supply units PW X023.

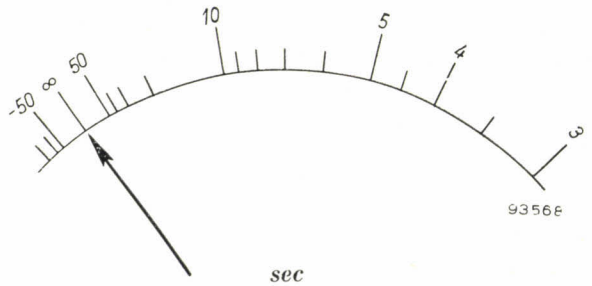


Fig. 10. Scale of period meter for high neutron flux. When the reactor is operating in the stationary state, the meter needle shows infinity. A positive deflection indicates increasing power, a negative deflection decreasing power. Audible and visible alarm signals are given if T drops below 10 sec; "fast scram" results automatically if T drops below 3 sec.

Linear measuring channel for high neutron flux

Owing to the wide range they cover, the logarithmic channels are especially suitable for use when putting the reactor into operation and for obtaining a rough impression of the value of the neutron flux at a given moment; they are also needed for measur-

ing the reactor period. A linear channel, however, which in this reactor has 18 measuring ranges, is used for accurate measurement, and also as the basis of the automatic control.

The neutron detector is a compensated ionization chamber of the same type as that in the logarithmic channel for high neutron flux. The current from this detector is passed through a high resistance. To cover the range of $6\frac{1}{2}$ decades, 18 such resistors with values rising in the ratio 1, 2, 5, 10, ... (to a maximum of 10^9 ohms) are mounted together with a high-insulation switch in a magnetically and electrically screening housing. The switch can be operated from the control desk.

The voltage across the resistor is measured with the aid of a vibrating capacitor [5]) and an amplifier (unit PR 7500) combined with an automatic recording potentiometer (type PR 2201). The linear scale of this "lin. N recorder" runs from zero to 10.0 (fig. 7).

Safety channel

As remarked, the safety system of the reactor is kept as far as possible separate from the other measuring apparatus for the sake of maximum reliability, the requirement being that the safety rods should fall not only when something is amiss in the reactor but also if any fault whatsoever should develop in the safety installation itself. Since inconvenience and loss of time are caused every time the rods fall, measures must be taken on the other hand to reduce to a minimum the chance of the rods being dropped without any real necessity. In the design of the safety installation, therefore, the aim must be to use as few components as possible, and these must be robust and of very high quality. Furthermore, the wiring system needs careful consideration.

The link between a very robust ionization chamber and the relay contacts through which the energizing current flows to the electromagnets from which the safety rods are suspended, is formed by the *safety amplifier*. This will be described in detail in Part II. In the event of too high neutron flux or internal disturbances, this device stops the reactor in the shortest possible time ("fast scram"). The same happens if the reactor period is too short: a relay in the "log. N and period" unit PW X081 interrupts the magnet currents if T drops to less than 3 seconds. This is done via a relay contact in the D.C. circuit of the electromagnets and also, as an extra precaution, via a relay contact on the A.C. side of the rectifier which supplies the magnet current.

There are further contacts incorporated on the A.C. side, their purpose also being to stop the reactor if certain troubles develop. Because of the energy accumulated in the rectifier smoothing circuit, the opening of an A.C. contact causes the magnet current to drop less rapidly to zero than the opening of a D.C. contact. This "slow scram", as it is called, is sufficient for certain less urgent cases (see *Table I*) and does not complicate the fast-scram line.

Thirdly, there are several relays which serve not to stop the reactor but only to set off a visible and sometimes also an audible alarm; see *Table II*. (In Delft the number of cases calling for "slow scram" and alarm signals will be extended.)

On the right-hand operating panel (fig. 7) the "safe" and "unsafe" signals are given by the burning of green and red pilot lamps respectively. This duplicated signalling system is again an extra safety precaution: if a red pilot lamp should fail to light up, the fact that the associated green lamp is not burning is a warning that something is wrong. The top row of pilot lamps relates to the "scram alarms", the second row to cases 1) to 5) in Table II (in these cases a buzzer signal is sounded at the same time). In cases 6), 7) and 8) the only indication is given by the pilot lamps in the bottom row.

Table I. Less urgent cases in which the reactor is stopped ("slow scrams").

Case	Shut-down device
1) Shut-down by hand	Push-button on central control panel (fig. 7) and at several points around the reactor.
2) Too low or no voltage on the compensated ionization chambers	Relay in each stabilized voltage source.
3) Excessive gamma radiation	Four meter relays in the gamma indicator (see Part II).
4) Reading on "log. C.R. meter": a) < 2 (due to fault in meter or because fission chamber is too high)	Slide contacts and relay in "log. C.R. recorder".
b) $> 10^4$, while reading on "log. N" $< 2\times10^{-4}$% (due to fault in "log. N meter", or because fission chamber is too low)	Slide contacts and relay in "log. C.R. recorder" and in "log. N recorder".
5) Reading on "log. N" $>$ 150 % (neutron flux too high)	Slide contacts and relay in "log. N recorder".
6) One or more of the safety rods not attached to their electromagnets	Contacts on the magnets, which are closed only if rods are properly attached.

[5]) See, for example, Philips tech. Rev. 7, 28, 1942.

Table II. Cases in which the reactor is not stopped, but an audible and visible alarm signal is given (cases 1 to 5) or only a visible signal (cases 6, 7 and 8).

Case	Device
1) $T < 10$ sec (pre-warning before period scram at $T < 3$ sec)	Relay in the period amplifier (see Part II).
2) Unusually large deviation of servomechanism from equilibrium position	Relay in control channel.
3) Power exceeds 110 % of nominal value (pre-warning before high-flux scram at 150 %)	As (5) in Table I.
4) Shim rods need adjustment because control rod has moved beyond its range	Contact in the path of control rod.
5) Defects in power supply for servomotors	Relay in control channel.
6) Water temperature near ion exchanger too high	Thermo-switch in ion exchanger.
7) Water temperature in the tank too high	Thermo-switch in the tank.
8) Conductivity of water too high	Relay in a bridge that measures the water conductivity.

Other measurements

Temperature measurements

Although there could scarcely be any appreciable rise in the temperature of such a large volume of water (50 m³) at the low maximum power of 10 kW at which the reactor operated at the Exhibition, the temperature was nevertheless measured at several points. This measurement will be of more importance when the power is raised to 100 kW, since the effective multiplication factor k in a reactor is closely dependent on the temperature (having a negative temperature coefficient).

The measurement is carried out by means of resistance thermometers, consisting of a platinum wire (100 ohms) in a sheath of stainless steel which is soldered water-tight to the lead sheath of the connecting cable. The platinum resistance element forms an arm of a bridge circuit, the output voltage of which, after amplification, deflects the needle of a meter calibrated in °C. This meter can be seen below the clock in fig. 7. At the right of the meter is a knob for switching over to different thermometer elements; at the left is unit PR X010 with the bridge circuit and amplifier.

Rod-position indicators

It is desirable to be able to read on the control desk the position of the control rod and safety rods. This will also be of use in experiments concerning the sensitivity of the control process at different positions of the rods.

The position indication is effected by the mechanical coupling of each rod to the slide contact of an accurate helical potentiometer ("Helipot") which is connected to a source of constant voltage. The voltage between the slide contact and a point of reference (corresponding to the lowest position of the rod) is measured with a voltmeter. The three voltmeters that indicate the positions of the safety rods can be seen top left on the middle control desk in fig. 7.

These meters do not give an accurate enough reading for all cases, and certainly not for the control rod. For exact position indication, therefore, an electronic voltmeter with digital display is used (product of Non-Linear Systems Inc.) whose reading is determined by separate potentiometers coupled to the control rod and to the safety rods. This meter, which is to be seen immediately on the right of the three voltmeters in fig. 7, gives a direct reading to four figures. As a rule the meter is employed for continuously indicating the position of the control rod, but by means of a push-button and a switch it can also indicate the position of each of the safety rods and that of the fission chamber for low neutron flux.

Gamma indicator

A special ionization chamber with electrometer tube is mounted near the ion exchanger and two others at points above the surface of the water. The dose rate of the gamma radiation at each position produces a reading on a logarithmically graduated meter; the latter can be seen on the left in fig. 3. Further particulars will be given in Part II.

Automatic control of the reactor

The design of the nuclear reactor here discussed provides for the automatic control of the neutron flux (and hence of the power output) once this has been set to the required value by hand. The reactor is not started up automatically, as some high-power reactors are, but by the following procedure.

First of all, the safety rods are raised "manually", that is to say with the hand switch that controls the servomotors. While this is being done, the neutron flux and period meters must be watched

to see that the rods are not raised too quickly. Only when they have reached a position where k is almost 1 can the operator begin to raise, very carefully, the control rod. Initially, only the meter for low neutron flux gives a reading, as a result of the activity of the artificial neutron source, which supplies the first neutrons. As the control rod gradually rises, fewer neutrons are absorbed and the low-flux period meter now gives a reading.

With the raising of the control rod the neutron multiplication factor k, which initially was only about 0.9, gradually increases until it is finally greater than 1. The chain reaction in the uranium now sets in, and at the same time the neutron flux spontaneously rises, i.e. without the control rod being moved any further. This produces a reading on the meter for high neutron flux. By fractional displacements of the control rod the rate at which the flux increases can be slightly altered, to keep it down to, say, 5% per second (period longer than 20 seconds).

In this way the flux can be brought to the required value in about 10 minutes. Once this is reached, the control rod is lowered to a position where k is exactly 1 (the period meter then shows infinity and the counting-rate meter gives a constant reading). In practice it is not possible to maintain this state exactly, but it can be very closely approached, so that slight adjustments of the control rod are needed only after long intervals. (However, the instruments would have to be kept under constant observation.)

This monotonous work can better be entrusted to an automatic control mechanism, which functions as follows.

The recording pen of the instrument which registers the neutron flux on a linear scale is mechanically coupled with the slide contact of a potentiometer. This forms a bridge circuit with a second potentiometer, which can be manually adjusted to accord with the power required. The bridge output voltage ε is thus a measure of the difference between the prevailing neutron flux and that corresponding to the required power. The principle of the control is that the voltage ε is fed to an amplifier, which produces an alternating voltage proportional to ε; this alternating voltage actuates one phase of a two-phase servomotor, which causes the control rod to move in the appropriate direction.

In reality the circuit of the amplifier designed for this purpose (type PR X083) is rather more complicated. Apart from a negative feedback which is linearly dependent on the position of the control rod, the amplifier contains a differentiating and an integrating element. The relation between the shift x of the control rod, the potential difference ε and the time t is accordingly:

$$x = A\left[\varepsilon + \tau_d \frac{d\varepsilon}{dt} + \frac{1}{\tau_i}\int \varepsilon dt\right],$$

in which the factor A and the time constants τ_d and τ_i are variable quantities. The differential term makes it possible for the final state to be reached more rapidly without the risk of instability, and the integral term allows the required state to be reached exactly instead of just approximately [6]).

The second potentiometer mentioned above, being another "Helipot" wire potentiometer, enables the required power to be adjusted with considerable precision. Ten revolutions of the knob are needed to move the slide contact over the whole resistance wire. The knob (marked "Power Demand") is on the right of the central control desk (fig. 7).

Since they differ in certain respects from the instruments commonly used for the same purpose, the logarithmic neutron flux and period meter for high flux, the gamma indicator and the safety amplifier will be dealt with more in detail in Part II.

Appendix: Fuel consumption and reaction products of a nuclear reactor

The fact that a nuclear reactor consumes extremely little "fuel" appears from the figures given below.

One gram-atom of U^{235}, i.e. 235 grams, contains 6.02×10^{23} atoms (Avogadro's number). Therefore one gram U^{235} contains $6.02\times10^{23}/235 = 2.56\times10^{21}$ U^{235} nuclei, and upon fission these yield $2.56\times10^{21}\times3.2\times10^{-11} = 8.2\times10^{10}$ joules $= 8.2\times10^{7}/60^2$ kWh $= 2.28\times10^{4}$ kWh. The generation of, say, 100 kW (in the form of heat) over a period of 24 hours therefore requires only $2400/(2.28\times10^4) \approx 0.1$ gram U^{235}. To generate the same amount of heat by complete combustion of coal, 300 kg of coal would be needed.

Of the substances produced in a nuclear reactor, the first to be mentioned are the *fission products*, i.e. the fragments into which U^{235} nuclei have split and whatever may come forth from these fragments. A U^{235} nucleus can split into two parts in more than 45 ways, over 90 different fission fragments having been identified (including the elements 43 and 61 — technetium and promethium respectively — which have never been found in nature). Owing to a surplus of neutrons, most fission fragments are not stable. Some of these neutrons are transformed in the nucleus into protons, during which process the nucleus emits an electron (beta emission) and possibly a gamma quantum, while others are liberated, after a beta emission, as "delayed" neutrons (see above). Often several beta emissions are needed before a radioactive fission fragment can reach nuclear stability. Two examples of such a disintegration in four stages accompanied by beta emission are:

$$_{54}Xe^{140} \xrightarrow[\beta]{16s} {}_{55}Cs^{140} \xrightarrow[\beta]{66s} {}_{56}Ba^{140} \xrightarrow[\beta]{12.8\,d} {}_{57}La^{140} \xrightarrow[\beta]{40h} {}_{58}Ce^{140}\ \text{(stable)},$$

[6]) See, for example, H. J. Roosdorp, On the regulation of industrial processes, Philips tech. Rev. **12**, 221-227, 1950/51.

$$_{36}Kr^{90} \xrightarrow[\beta]{33s} {}_{37}Rb^{90} \xrightarrow[\beta]{2.7\ min} {}_{38}Sr^{90} \xrightarrow[\beta]{19.9\ y} {}_{39}Y^{90} \xrightarrow[\beta]{61h} {}_{40}Zr^{90} \quad \text{(stable)}.$$

An example of a case in which a delayed neutron is liberated is:

$$_{35}Br^{87} \xrightarrow[\beta]{55.6\ s} {}_{36}Kr^{87} \xrightarrow[n]{0} {}_{36}Kr^{86} \text{ (stable)}.$$

The half-life of these substances is mentioned in the examples. In some cases the half-life is so long (e.g. 19.9 years for Sr^{90}, see the second example) that the question of where the fission products are to be disposed of requires careful consideration.

Another series of transformations taking place in a uranium reactor begins with the absorption of a neutron of 6.5 eV by a U^{238} nucleus (resonance). This gives rise to $_{92}U^{239}$, which, with two beta emissions, transforms into plutonium ($_{94}Pu^{239}$), with neptunium ($_{93}Np^{239}$) as intermediate product. Plutonium can be separated by a chemical process from the mixtures of reaction products, and put to further use as nuclear fuel, Pu^{239} nuclei being fissionable by neutron capture.

Finally, it may be noted that, because of the low power of the reactor shown at the Exhibition "Het Atoom", the fission products constituted no problem and there was no significant production of plutonium.

Summary. The monitoring, control and safety equipment described relates to the experimental nuclear reactor (of American manufacture) which was exhibited in operation at the 1957 exhibition "Het Atoom" in Amsterdam, and which is shortly to be transferred to Delft. A brief description is first given of the operation of a nuclear reactor. The equipment contains the following channels for the measurement of the neutron flux: a low-flux channel (with logarithmic scale), two high-flux channels (one with logarithmic and one with linear scale) and two identical independent safety channels. The two logarithmic channels each contain a period meter, which indicates the "period" (time constant) of the nuclear reactor directly in seconds. The installation includes devices for measuring the temperature of the water and for indicating the position of the control and safety rods, a gamma-ray indicator and an automatic control system. An appendix gives a few details concerning the fuel consumption and the reaction products of a nuclear reactor.

A RADAR SONDE SYSTEM FOR UPPER AIR MEASUREMENTS

621.396.969: 551.508.8

The measurement of upper air conditions is an important branch of meteorology because information is required for research, for weather forecasting and for aircraft operation. Although the first soundings were made in the 18th century, using kites and manned balloons, the systematic measurement of air conditions at great heights has become possible only since the development of radio and radar. The development of radio and radar sonde equipment is therefore one of the most significant advances in meteorological instrumentation.

In the following note, a short description will be given of a radar sonde system which has been developed by the Mullard Research Laboratories, Salfords (England), in conjunction with the Royal Radar Establishment, Malvern, for measurements of wind speed, wind direction, temperature, pressure and humidity at heights up to at least 24 km (80 000 ft) [1]). The equipment is automatic in operation and provides continuous records at a ground station of the wind vector as a function of height. Temperature, pressure and humidity are measured and recorded at the ground station at approximately 100 m intervals of the height of the sonde.

For each sounding, an airborne unit (which in a large percentage of soundings is not recovered) is carried into the upper atmosphere by a free, hydrogen-filled balloon. This unit (*fig. 1*) contains a small combined transmitter and receiver with auxiliary equipment and instruments for measuring temperature, pressure and humidity. The transmitter allows the telemetering of these data during the whole ascent, until the bursting of the balloon, which may occur after it has reached a height greater than 80 000 feet and travelled a distance of over 100

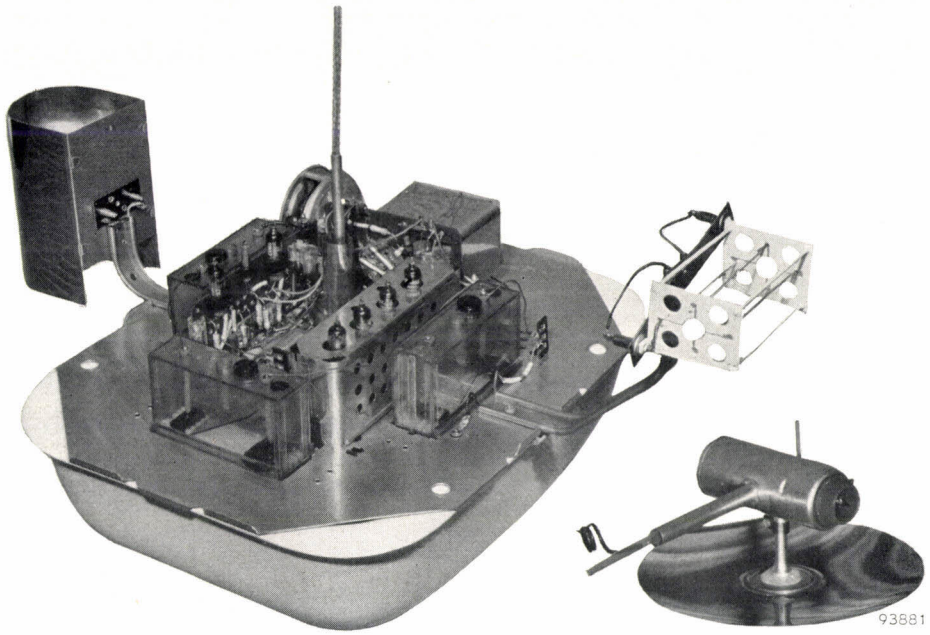


Fig. 1. Airborne unit of the radar sonde (cover removed). The humidity and temperature measuring elements are carried on external arms; the main part of the unit is enclosed in a thermally insulating case. The aerial, receiver, telemetering circuit, switch, barometer and batteries are seen mounted on the chassis plate. The 10 cm wave transmitter and aerial assembly, which are mounted underneath the chassis plate, are shown separately on the right-hand side of the photograph.

The airborne unit is suspended from a hydrogen-filled balloon which is approximately 2 m in diameter at the ground and lifts the unit at 360 m per minute (1200 ft/min). A parachute is inserted in the suspension to allow a gentle descent after the balloon has burst (at a height greater than 80 000 feet).

[1]) See also: F. E. Jones, J. E. N. Hooper and N. L. Alder, The radar sonde system for the measurement of upper wind and air data, Proc. Instn. El. Engrs. **98**, II, 461-469, 1951.
A. L. Maidens, Sounding the upper atmosphere, Discovery **18**, 156-160, 1957 (No. 4).

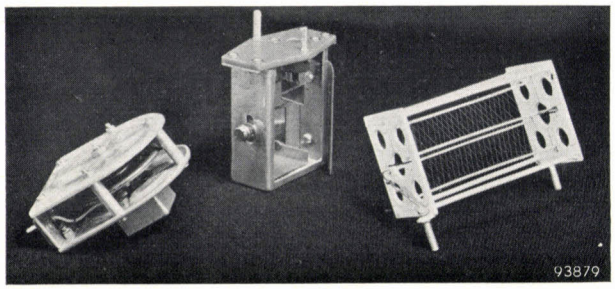


Fig. 2. *Left*: the aneroid barometer; *right*: the fine-wire resistance thermometer; *centre*: the skin hygrometer. The aneroid and hygrometer are conventional elements, but the thermometer is a considerable improvement over earlier types. It is constructed with fine, coiled-coil, tungsten wire, has a very short time constant and a very small radiation error at high altitude.

miles from the release point. Tracking the flight of the balloon by radar at the ground station provides the data necessary for determining the wind vector and height at each moment. Conventional radar methods, which depend on a pulse signal transmitted by the ground station being *reflected* by the target, would involve a very high transmitting power in order to enable the balloon to be tracked during its whole flight. Therefore in this radar sonde a system known as "secondary radar" has been employed. The airborne transmitter is made to send a pulse (or in fact 2 pulses, see later) every time the airborne receiver picks up a pulse signal transmitted by the ground station. The time delay measured at the ground station between the departure of the "interrogating pulse" and the arrival of the "responding pulse" is a direct measure of the range of the balloon; the azimuth and elevation angles of the transmitting aerial, which is automatically kept aligned on the balloon, convey the information concerning its direction.

The interrogating signal has a frequency of 152.5 Mc/s and is transmitted at a peak power of 50 kW with a pulse repetition rate of 404 pulses per second. The responding signal of the airborne transmitter has a frequency of 2850 Mc/s and is radiated at a peak power of 0.03 kW.

For the sensing of pressure, humidity and temperature, the airborne unit carries an aneroid barometer, a goldbeater's skin hygrometer and a resistance thermometer (*fig. 2*). The changes in these elements under varying atmospheric conditions are converted into voltage variations which, in turn, control an "encoding" circuit. This circuit is triggered by each of the interrogating pulses, derived from the receiver, and generates a pulse of its own with a time delay dependent on the control voltage. The airborne transmitter is operated both by this delayed pulse and by the interrogating pulse and thus transmits back to the ground a pair of pulses for each interrogating pulse, the time delay between the two pulses of a pair representing the reading of the meteorological instrument. The control voltages of the three instruments are brought into operation in sequence by a motor-driven switch, so that each instrument reading is encoded once every 17 seconds.

The airborne centimetric transmitter contains a triode transmitting valve, controlled by a blocking-oscillator modulator and mounted in a coaxial cavity designed to facilitate large-scale production. The transmitter aerial is an unipole with counterpoise (fig. 1). The power supply of the sonde comprises three primary cells and a vibrator.

The ground equipment is divisible functionally and physically into two groups. The *radar* group is concerned with the interrogation of the airborne unit, reception of the return signal and determination of the basic positional data (azimuth angle, elevation angle and slant range of the balloon as a function of time). The *computer* group continuously translates the positional data into wind speed and direction, decodes the telemetering signal and records all the results in graphical and printed form.

The radar equipment is illustrated in *figs 3* and *4*. The transmitting aerial (a Yagi array) and the

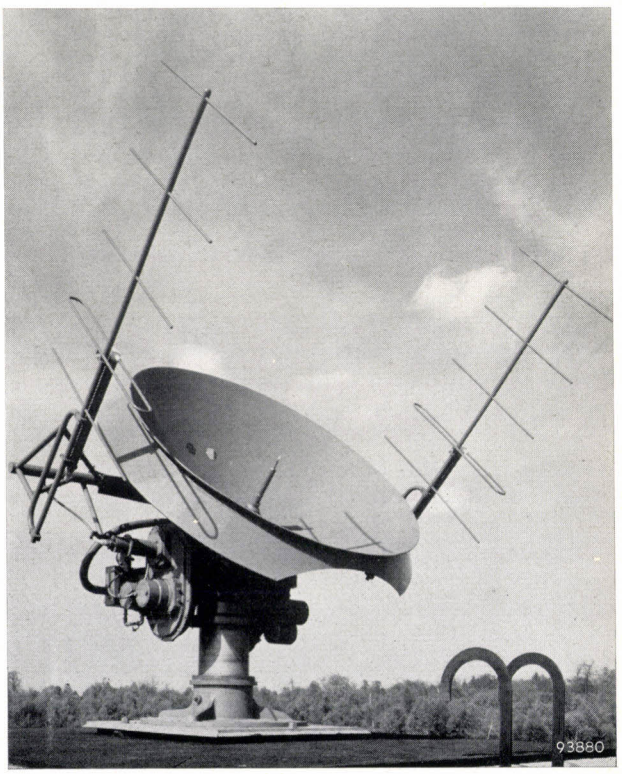


Fig. 3. The Yagi array transmitting the 50 kW interrogating pulses of 2 m waves is mounted on a common pedestal with the receiving aerial, a 5 ft diameter parabolic reflector with nutating dipole. They are automatically kept aligned on the balloon.

receiving aerial (a parabolic reflector) are mounted on a common pedestal driven by a motor servosystem. The receiving aerial is fitted with a nutating dipole which produces a conical scan of the narrow aerial beam and superimposes a modulation on the received signal when the aerial is incorrectly aligned. This error signal steers the servosystem so as to give accurate automatic following of the balloon in direction. The microwave receiver, which is equipped with automatic gain and frequency control, amplifies the signal from the balloon and delivers it to the automatic ranging system and the telemetering system. The ranging system is of a novel type in which the rotation of a motor-driven, of a mile to be measured; full tenths of a mile are determined by counting the number of oscillations of a crystal calibrator occurring between transmitted and responding pulse. The calibrator controls the pulse repetition frequency of the transmitter and oscillates with a time period exactly corresponding to one tenth of a nautical mile in radar range. The received signal is monitored and readings of range R, azimuth Θ and elevation E are provided by a display unit (fig. 4). This unit is also fitted with the manual controls for range and aerial position which are used at the beginning of a flight before the equipment is finally switched over to fully automatic operation.

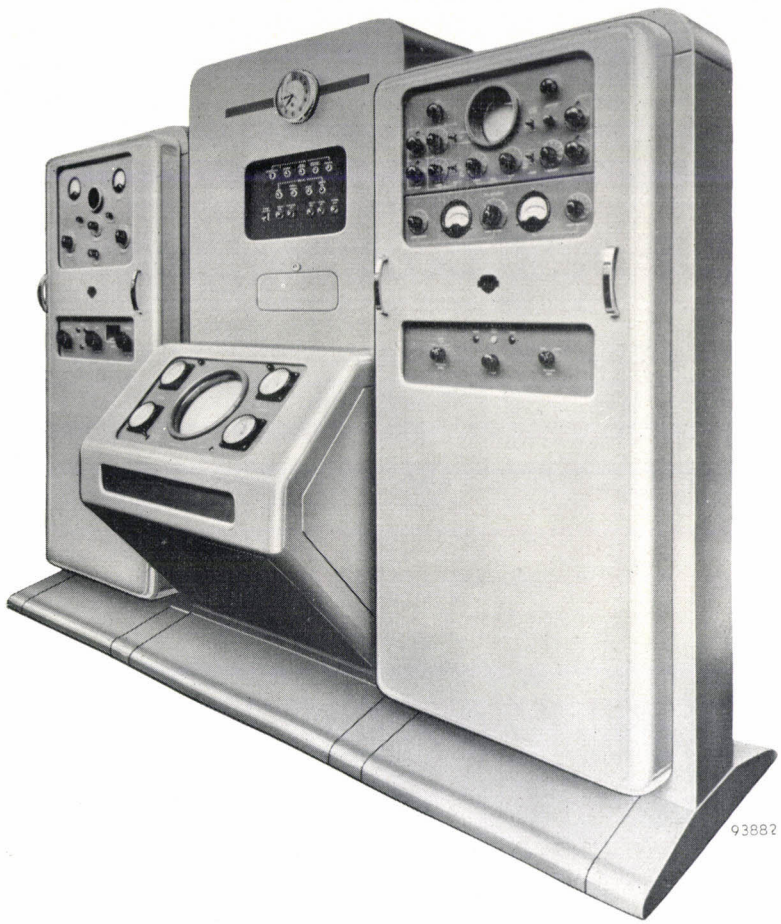


Fig. 4. The radar console. The left-hand unit contains the 50 kW transmitter, the right-hand unit the 10 cm wave receiver and the automatic ranging and aerial-alignment equipment. The centre unit is the display and control system.

phase-shifting transformer is related to the change in distance of the balloon, so that the speed of rotation is an accurate measure of the radial velocity of the balloon (radial wind component). A rotation of 360° is equivalent to a change in distance of one tenth of a nautical mile. The phase-shifting transformer, therefore, permits fractions of a tenth

The continuous computation of the wind vector (which is parallel to the surface of the earth) is effected from the tangential and radial wind components, given by the equations (in spherical polar coordinates):

$$V_T = R\,(\mathrm{d}\Theta/\mathrm{d}t)\cos E,$$

$$V_R = (\mathrm{d}R/\mathrm{d}t)\cos E - R(\mathrm{d}E/\mathrm{d}t)\sin E.$$

An important feature of these equations is the occurrence of the derivatives of (slant) range R, azimuth Θ and elevation E with respect to time t, i.e. the radial and angular velocities. The auto-ranging and auto-alignment circuits are so designed as to provide signals directly proportional to these velocities. The wind computer which determines the vector resultant of V_T and V_R is illustrated in *fig. 5*. It is an analogue computer employing precision potentiometers for multiplication and magslips (synchros) for trigonometrical computation. The vector amplitude and direction are derived from V_T and V_R in a "triangle-solver" magslip whose rotor shaft assumes a position corresponding at every moment to the wind direction. A 4 ft diameter recording table is directly coupled to this shaft and a recording pen is driven at a constant speed in a radial direction across the table. The pen thus traces out a continuous graph of instantaneous wind direction against time. A voltage proportional to the vector amplitude, i.e. the wind speed, is induced in the magslip rotor coil and is recorded by a conventional recording meter. The height of the balloon, corrected for earth curvature, is computed by similar analogue methods and recorded as a function of time.

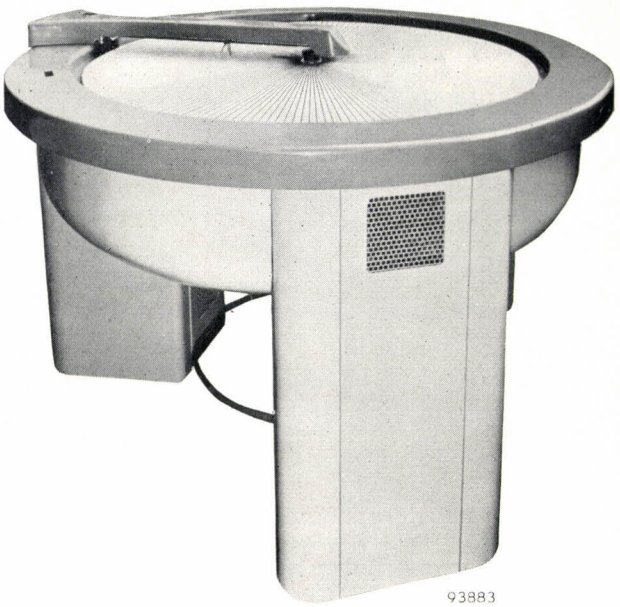


Fig. 5. The wind computer. A magslip turns the rotatable 4 ft diameter table so that at every instant it has a position corresponding to the computed wind direction. A recording pen moving at constant speed along the fixed radial arm traces out a graph of wind direction as a function of time (cf. fig. 7).

The ground telemetering equipment, shown in *fig. 6*, is designed to identify the coded signals from the balloon, to measure accurately the time delay between the ranging and telemetering pulses and to encode the results in a form suitable for operating a standard teleprinter. The time delay is measured by allowing crystal-controlled timing pulses, with a one microsecond separation, to pass through a gating circuit during a period initiated by the ranging pulse and terminated by the telemetering pulse. The total number of timing pulses corresponding to five hundred pairs of signal pulses is counted by an electronic counter and the average delay is transferred to the teleprinter encoder. The figures are printed by the teleprinter in five columns, each of four digits, corresponding to the three meteorological parameters and two reference signals which are also transmitted by the airborne unit. Coding sig-

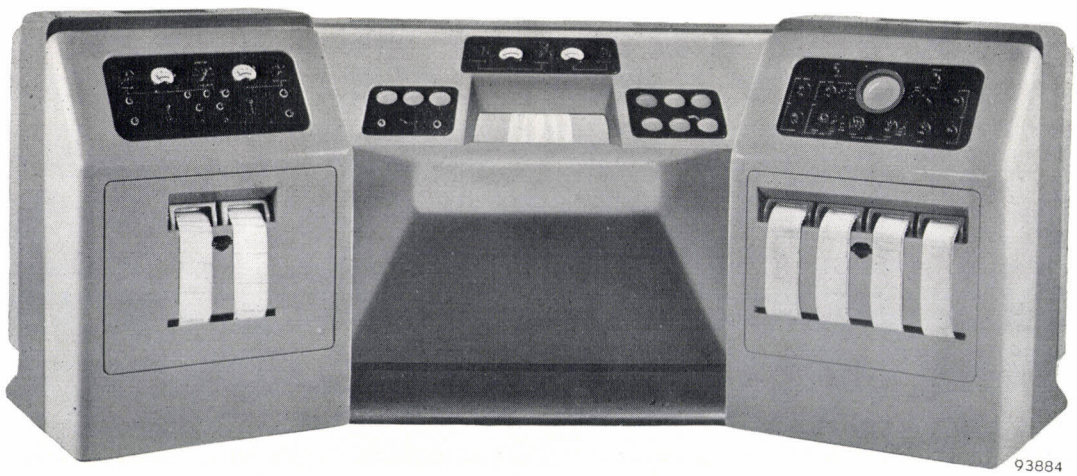


Fig. 6. The telemetering console. The left-hand unit contains wind-speed and height recorders, the right-hand unit encoders and temperature, pressure and humidity recorders and the centre unit a digital computer and teleprinter.

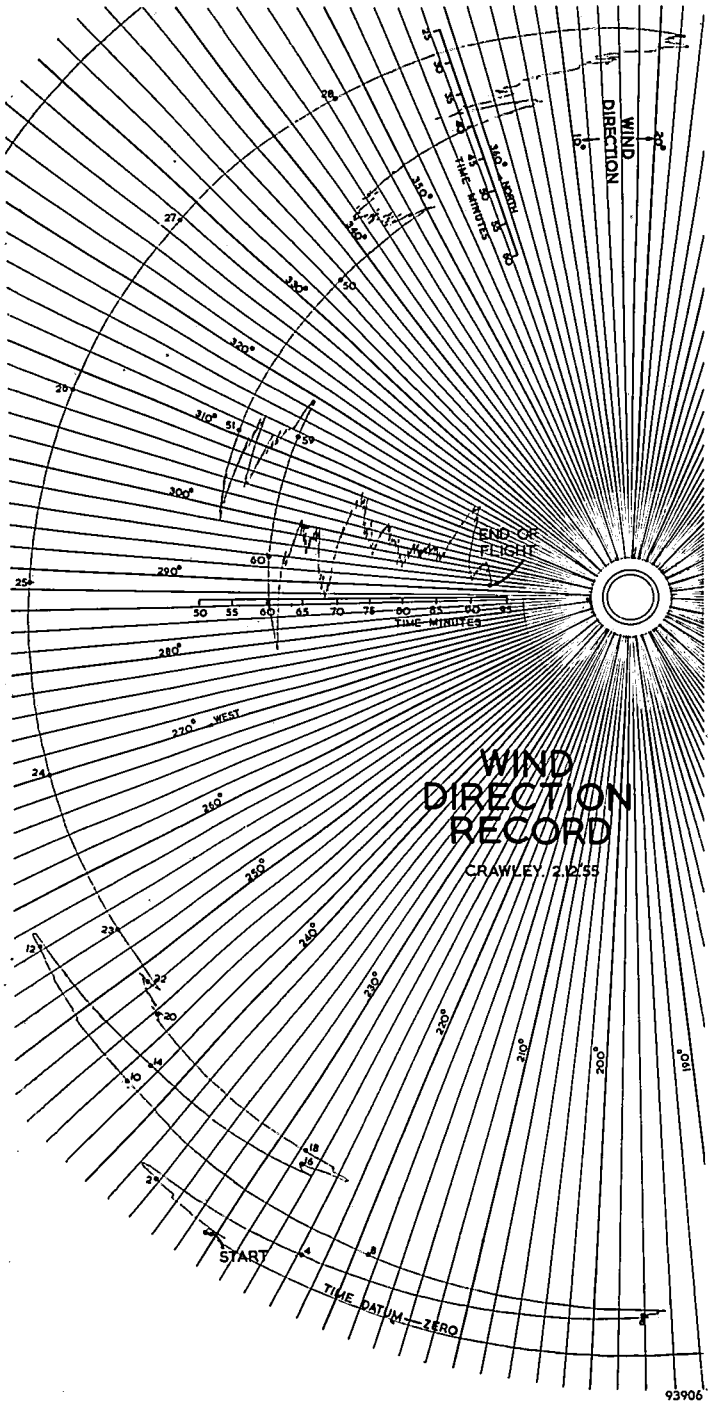


Fig. 7

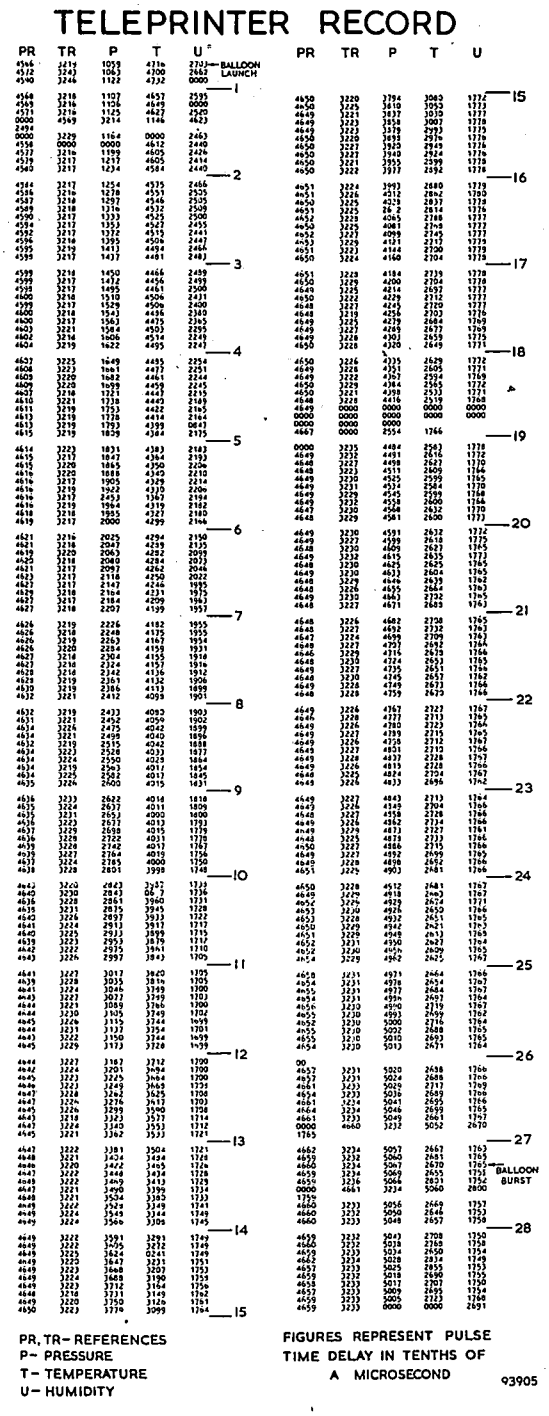


Fig. 8

Fig. 7. A typical wind direction record obtained from the radar sonde system.
Fig. 8. Typical teleprinter record obtained from the radar sonde system.

nals interposed in the airborne unit switching sequence are used to ensure that the results are printed in the correct columns.

Some typical flight records, which are self-explanatory, are shown in *figures 7, 8* and *9*. The equipment is designed to measure wind speed with an error not exceeding 5 km/hr. The telemetering channels are designed for an error of the order 0.1%, but at present the meteorological instruments themselves contribute somewhat larger errors.

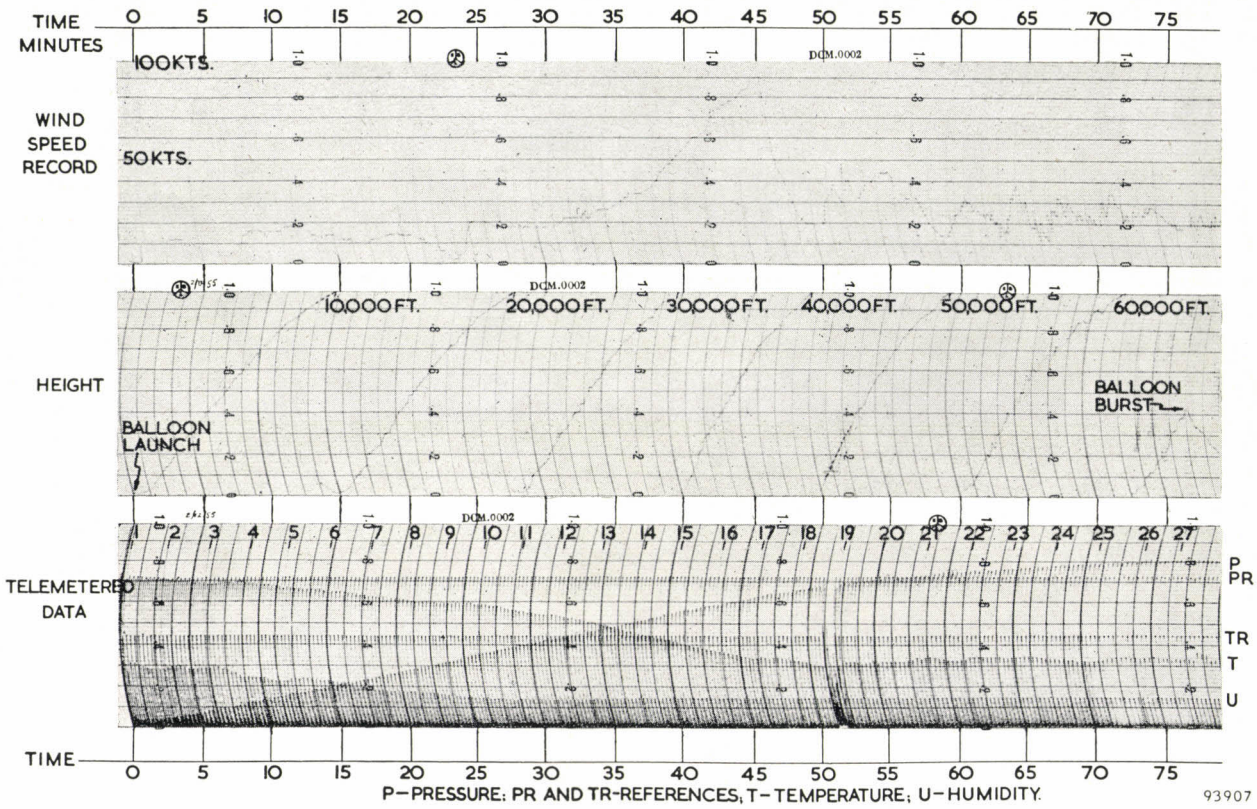


Fig. 9. Example of wind-speed, height and telemetering records obtained from the radar sonde system.

The prototype equipment is at present undergoing flight trials to assess its suitability for routine soundings and research.

The authors are indebted to the Director of the Mullard Research Laboratories, the Ministry of Supply and the Meteorological Office for permission to publish this note.

N. E. GODDARD*) and H. A. DELL*).

*) Mullard Research Laboratories, Salfords (England).

DOSIMETRY OF THE VERY WEAK X-RADIATION GENERATED IN TELEVISION RECEIVERS AND X-RAY DIFFRACTION APPARATUS

621.386.82: 621.397.62

With many kinds of apparatus in which electron tubes are operated at high voltages, the possible presence of very weak, soft X-radiation (i.e. X-radiation of low penetrating power) in the neighbourhood must be given due consideration. The question arises in such apparatus as X-ray diffraction equipment, formerly found only in research laboratories but now widely used industrially in inspection and production departments. Television receivers, too, can emit X-rays, for in modern large-screen sets the accelerating voltage for the electron beam in the picture tube can be as high as 20 kV at a beam current up to 0.2 mA. Fortunately this X-radiation is so soft that it can be attenuated to harmless intensities by simple means, for example by the 6 mm thick glass shield over the viewing screen, which is needed in any case as protection against implosion of the tube. Nevertheless it is necessary during manufacture to make dosimetric tests to ensure that the intensity of the radiation at the outside of the set in no way endangers the user.

The International Commission for Radiological Protection [1]) has recommended that the dose rate at the outside surface of home television receivers should not exceed 2 milliröntgens per hour (2 mr/h or 0.6 μr/sec). In the future the limit will probably be set lower still, in order to preclude all possibility of genetic hazards. This possible lowering of the limit has been allowed for in the design of the dosimetric apparatus described in this article.

In the case of X-ray diffraction apparatus, the requirements regarding the maximum permissible dose rate outside the useful X-ray beam are less severe, such apparatus being used by so very small a section of the population that genetic considerations do not arise and only the norms for individual protection need be applied. For the industrial use of X-ray equipment the recommended maximum permissible dose rate outside the useful beam is 6 mr/h.

The dose rate of very weak, soft X-radiation, such as that from television receivers, can be checked with Geiger-Müller counter tubes fitted with a window of thin mica (approx. 15 μ). Such tubes are quite sensitive enough for the purpose and the mica window is sufficiently transparent to these soft rays [2]). It has also been found that if the tube has a suitable gas filling, the sensitivity varies only slightly with X-ray wavelength, for the accelerating voltages common in television picture tubes (10-20 kV) [3]). Another advantage of Geiger-Müller counters is their small effective surface (approx. 10 mm diameter). The reading on the meter (for example, a counting-rate meter [2])) then corresponds to the dose rate averaged over the effective surface of the detector. Geiger-Müller counters will therefore be much more sensitive to narrow beams of leakage radiation than the ionization chambers more generally used in dosimetry, which have an effective surface 10 to 100 times larger.

An important problem in radiation measurements is the *calibration* of the meter. This must be done with the aid of a standard dosemeter which has been so designed as to enable the radiation dose to be calculated from the meter reading and the geometrical and electrical data. The standard dosemeter generally used by us for soft radiation, an instrument based on a design by Taylor and Stoneburner [4]), is shown schematically in *fig. 1* [5]). The caption explains the measures taken to make the instrument satisfy the conditions implicit in the definition of the röntgen unit. The dose rate in röntgens per second at the diaphragm D is calculated by dividing the collector current, expressed in e.s.u./sec, by the mass of air, expressed in units of 1.293 mg, in a cylinder having a length equal to the distance between the planes AA' and BB' and having the diaphragm aperture as its base. (A correction needs to be made for the attenuation undergone by the X-rays between the diaphragm and the plane through the centre of the collector.)

These standard dosemeters were made specifically for calibrating the dosemeters used in X-ray therapy, where the dose rates to be measured range from 0.5 to 5 r/sec, or sometimes appreciably higher.

[1]) Recommendations of the International Commission for Radiological Protection, Brit. J. Radiology, Suppl. No. 6, 1955.

[2]) P. H. Dowling, C. E. Hendee, T. R. Kohler and W. Parrish, Counters for X-ray analysis, Philips tech. Rev. **18**, 262-275, 1956/57 (No. 9).

[3]) See also: T. R. Kohler and W. Parrish, Conversion of quantum counting rate to roentgens, Rev. sci. Instr. **27**, 705-706, 1956 (No. 9).

[4]) L. S. Taylor and C. F. Stoneburner, Radiology **23**, 22, 1934.

[5]) W. J. Oosterkamp and J. Proper, Acta radiologica **37**, 33, 1952.

The dosemeters used for protection purposes are required to measure very much lower dose rates. As we have seen, the permissible maximum for television receivers is 0.6 μr/sec, and in the future it will probably be considerably lower, e.g. 0.1 or 0.2 μr/sec. In order to measure such low dose rates with reasonable accuracy, the instruments used must be calibrated by a standard dosemeter having a sensitivity several million times greater than that of the one illustrated in fig. 1.

The sensitivity of an ionization dosemeter can be increased by increasing the sensitivity of the current measurement. The ionization current is measured either by passing it through a resistor and determining the voltage drop across the resistor with an electrometer, or by causing the current to charge or discharge a capacitor and measuring the change in the voltage on the capacitor in a specific time interval. By using a resistor of higher value, or a capacitor of lower capacitance, the sensitivity of the current measurement can be raised. It is obvious, however, that there is a limit to this.

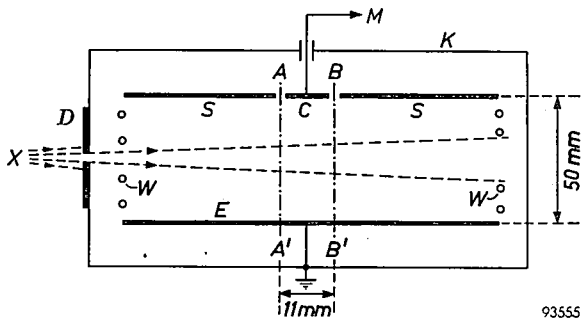


Fig. 1. Schematic cross-section of a standard ionization chamber for dosimetry of soft X-radiation. E and $C+S$ are plane parallel electrodes in a metal, air-filled box K. The X-ray beam X entering the chamber is limited by a round diaphragm D of lead or tungsten. The air in the path of the beam is ionized, chiefly by the fast electrons liberated in air upon primary ionization by the X-rays. These electrons having a fairly large range, penetrate the air around the X-ray beam and cause ionization there too, thus giving rise to a cloud of ions in and around the beam. The electrodes are far enough apart to allow this cloud to form unhindered. The middle section of one of the electrodes, the collector C, is insulated from the outer sections S, which act as a guard ring. Collector and guard ring have approximately the same potential. Owing to the guard ring and to the wires W at both ends of the electrode system, which are held at appropriate potentials, the electric field between AA' and BB' is very uniform, i.e. it has practically no axial component. As a result, provided the field strength is sufficient to draw the ions so rapidly to the electrodes that there is scarcely any chance of recombination, all ions of the same sign, formed between the planes AA' and BB', will travel to the collector. This satisfies the conditions implicit in the definition of the röntgen. The dose rate at the diaphragm, expressed in röntgens per second, is now equal to the collector current in e.s.u. per sec (measured by meter M) divided by the mass of air, expressed in units of 1.293 mg, in a cylinder of length equal to the inter-planar distance of AA' and BB' and of diameter equal to that of the diaphragm aperture. The volume of this cylinder is 0.08 cm³ in the chamber illustrated.

In the standard chamber described, the air volume in which the ionization is measured, and to which the ionization current is proportional, is 0.08 cm³. It can be readily calculated that at a dose rate of 0.15 μr/sec the ionization current in this chamber would be only about 10^{-8} e.s.u./sec $= 3\times10^{-18}$ A. Currents as small as this can no longer be measured by simple means with the required accuracy.

We circumvent this difficulty by calibrating the protection dosemeters with the intermediary of a *substandard chamber* having a volume of about 160 cm³, i.e. 2000 times larger than that of the standard chamber. With a suitably chosen dose rate of the radiation and an appropriate resistor or capacitor it is now possible to calibrate first the substandard chamber with respect to the standard chamber, and then, with the aid of a higher resistance, or a lower capacitance, to check the instruments to be calibrated against the substandard chamber.

In a chamber with such a large volume it is not easy to realize the conditions which must be fulfilled in order to calculate the number of röntgens directly from the geometrical and other data. This is not necessary, however, with the procedure described: all that is required of the substandard chamber is that the reading obtained should be readily reproducible, and that its sensitivity should not vary to any marked extent with the hardness of the X-radiation measured.

For reproducible results, all the ions formed in the chamber must be drawn by the field to the collector. In the first place, therefore, there must be no air currents in the ionization chamber that could carry away the ions, and secondly the electrical field strength must everywhere be high enough to ensure that the ions arrive at the collector before appreciable numbers have had a chance to recombine. The latter condition calls for the most uniform possible field, since it is also desirable to keep the voltage applied to the chamber electrodes as low as possible; an arrangement with flat, parallel electrodes is therefore used. The first condition, however, implies that the chamber should preferably be completely closed, or at all events should have only very small openings. The standard-chamber design sketched in fig. 1, in which the X-rays enter the chamber transversely to the direction of the electric field, is therefore less suitable where a large volume is to be irradiated, as in the sub-standard chamber. The design which we have adopted is illustrated in *fig. 2*. The chamber is in the form of a round box of methyl metacrylate, slightly more than 10 cm in diameter. The radiation enters the box

through a "cover" of thin fine-mesh nylon gauze, coated with graphite to make it conductive. The gauze, together with the graphite-coated bottom of the box, constitutes the electrode system; the holes in the gauze are very small so that air currents are avoided. At the same time the threads of the gauze are very thin and the aperture/thread ratio relatively large; hence the gauze gives rise to only slight attenuation of the very soft X-radiation to be measured.

In the standard chamber the air present in the measuring volume is bounded in the direction of the beam by air and the electrodes are outside the X-ray beam; this is not the case in the sub-standard chamber. Although the electrodes in the latter have been made as far as possible "air-equivalent", there is a possibility that not as many electrons will enter the measuring volume from the electrodes as are lost from this volume to the electrodes. It is for this reason that comparison with a standard chamber is desirable. In making this comparison it is necessary that the X-ray beam used should have a constant intensity over the large cross-section of the substandard chamber.

The results of comparing the sub-standard with the standard chamber are shown in *fig. 3* for X-radiation of varying hardness, the latter being

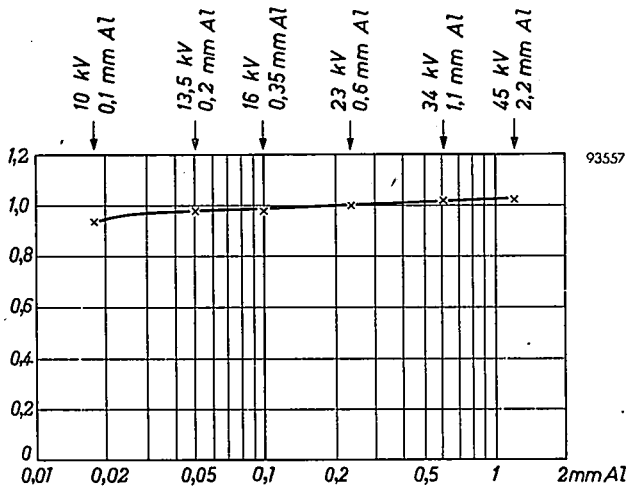


Fig. 3. Calibration curve of the sub-standard chamber illustrated in fig. 2. The ratio of the sensitivity of this chamber to that of the standard chamber in fig. 1, reduced to the same volume of air, is plotted as a function of the hardness of the X-radiation. The hardness is expressed by the thickness of that aluminium layer through which the rays must pass for the dose rate to be reduced by one half (half-value thickness in mm Al). For six different qualities of radiation, the corresponding tube voltages (kV) and filters (mm Al) are also indicated.

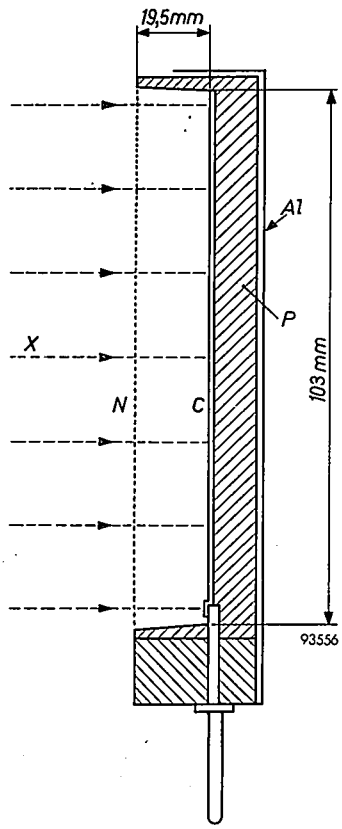


Fig. 2. Sub-standard chamber for dosimetry of extremely weak soft X-radiation. Owing to the large measuring volume (160 cm^3), a measurable collector current is obtained at dose rates as low as 0.1-0.2 μr/sec, which may be found at the outside surface of home television receivers. This sub-standard chamber is used to calibrate Geiger-Müller counters used as dosemeters for protection purposes. *C* graphite layer, acting as collector electrode. *N* electrode of nylon gauze, coated with graphite, through which X-rays enter the chamber. *P* methyl methacrylate box, the outer surface being lined with earthed aluminium foil for electrical screening.

expressed in terms of the half-value thickness in aluminium, or by the voltage on the X-ray tube and the filter in the beam.

A number of Geiger-Müller counters of different dimensions, window thicknesses and gas-fillings were calibrated with the sub-standard chamber at radiations of different hardness; the curves obtained are shown in *fig. 4*. For this purpose the counter tubes were connected to a count-rate meter unit (type PW 4041) which also provides the necessary high voltage supply.

Full deflection in the most sensitive range (300 pulses/min) corresponds to about 0.2 mr/h for counter MX 114, and to about 1.5 mr/h for counter MX 108. It should be recalled that the dose rates to be measured are of the order of magnitude of 0.5 to 2 mr/h. With this in mind and also because it is less dependent upon the hardness of the radiation than the others, the type MX 108 counter is the most suitable for the object in view.

With a Geiger-Müller counter calibrated in this way dose measurements were carried out on three types of television set (17 TX 144A-38, 21 TX 144A-00 and 17 TX 123U-00) operated at about 10% above their nominal anode voltage. At no point on the outside surface of the sets was the dose rate in excess of 0.5 mr/h. In earlier measurements on a similar television set the dose rates were found to be higher, up to 2 mr/h. In this old model, however, the X-rays were mainly caused by the rectifier tubes, type DY 86, in the extra-high tension generator, and only to a slight extent by the picture tube. By providing the rectifier tubes with a lead-rich glass envelope, this unwanted radiation was completely eliminated and thus the safety margin increased.

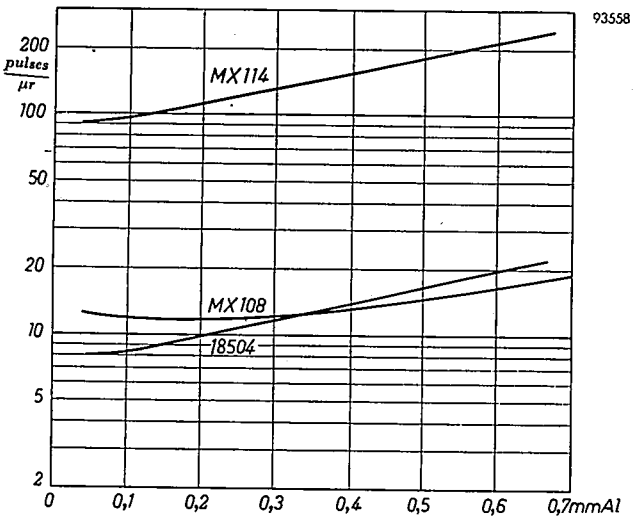


Fig. 4. Calibration curves of three different Geiger-Müller counter tubes, for X-radiation generated at voltages between 10 and 25 kV.
Type 18504: window thickness approx. 10 μ, window diameter 9 mm, gas-filling neon + argon + halogen quencher. The background counts during measurement are 20 pulses/min; useful measurements are still obtained if the signal is roughly equal to the background, i.e. in this case about 2 μr/min, or 0.1 mr/h (at 16 kV).
Type MX 108: window thickness approx. 15 μ, window diameter 17 mm, gas-filling neon + argon + halogen quencher. Background counts 35 pulses/min; lower limit of measurement approx. 0.2 mr/h.
Type MX 114: window thickness approx. 15 μ, window diameter 28 mm, gas-filling neon + argon + halogen quencher. Background counts 70 pulses/min; lower limit of measurement approx. 0.04 mr/h.

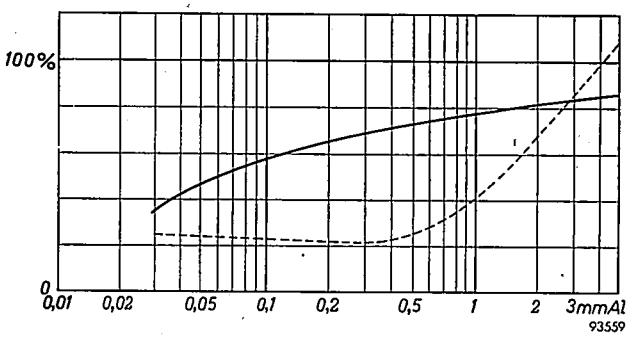


Fig. 5. Sensitivity of an ionization chamber (full curve, approaching 100% for very hard radiation) and of a Geiger-Müller counter type MX 108 (broken curve) as a function of the hardness of the X-radiation.

As already remarked, the recommended maximum permissible dose rate on the outside of X-ray diffraction apparatus is 6 mr/h. This can be measured with sufficient accuracy by ionization instruments (a much used instrument, for example, gives full deflection at 15 mr/h). Voltages up to 50 kV are used in X-ray diffraction apparatus, which means that harder as well as soft radiation is emitted. Above 20 kV the sensitivity of Geiger-Müller tubes increases sharply, whereas that of an ionization chamber remains fairly constant (*fig. 5*). This, too, argues in favour of using an ionization chamber. However, the latter has the drawback, mentioned earlier, of a large effective surface, and therefore it is best to use both instruments in conjunction with each other.

W. J. OOSTERKAMP, J. PROPER and J. J. F. de WIJK.

AN IMPROVED METHOD OF CLEATING

621.981.6

Cleating is a method of riveting widely used in the radio and electronics industry. The object is to effect a mechanical joint between two or more parts, for example between a bush, hub or nut (the cleat) and one or more plates or strips. The cleat has a cylindrical rim which, after it has been pushed through a hole in the plates or strips, is forced outwards by means of a special tool, thus completing the joint. The method is illustrated in *fig. 1*. Cleating can be carried out in two ways: by direct pressure or by

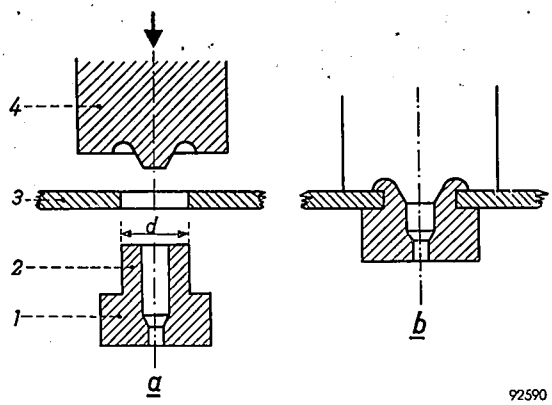


Fig. 1. *a*) *1* cleat, *2* "rim" of cleat, *3* plate with which joint is to be made, and *4* cleating tool for effecting the joint. *b*) Completed joint.

applying pressure with the tool spinning. The appropriate tools are represented in *fig. 2*. In the spinning method the tool rotates at a speed of several hundred revolutions per minute. This requires less pressure than the direct pressure method, which is an important consideration with brittle material (ceramics, plastics), but it takes somewhat longer to effect the joint.

The requirements of a cleated joint are concerned with mechanical strength and appearance. In the process of cleating there is a danger that both the strength and the appearance of the joint will suffer owing to the occurrence of cracks in the rim. (Corrosion may be another cause, but this will not be considered here.)

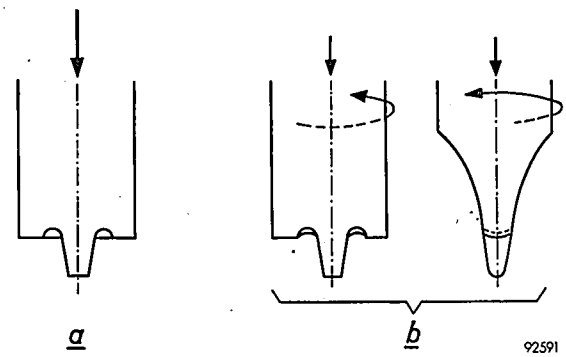


Fig. 2. *a*) Tool for direct pressure method (solid of revolution). *b*) Tool for spinning method (front and side views).

The cleats are machined from steel, brass or aluminium bar, which are free-cutting only after prior hardening. Hardening can be effected by cold-working, but this in turn reduces ductility and thus increases the likelihood of the rim cracking. For this reason the cleats are subjected, after machining, to an annealing treatment to eliminate the effects of cold-working. It often happens that the annealing process (during which the internal stresses are relieved) gives rise to deformation. If accurate dimensions are necessary, the cleats will therefore have to undergo further mechanical treatment, such as grinding or reaming. The costs involved in these different treatments have led the Metallurgical Laboratory of the Radio-set Factory at Eindhoven, to seek a more efficient solution.

In view of the extent to which the rim of the cleat is plastically deformed during the making of the joint, the occurrence of cracks is not surprising. In the case of a rim of diameter $d = 4$ mm (fig. 1*a*), a radial compression of about 15% occurs at point A (*fig. 3*) and a tangential elongation of about 25%; at point B there are radial and tangential elongations of as much as 70% and 100%, respectively. The fact that the metal can withstand such (by normal standards) large plastic deformation is due to the introduction during the cleating process of compressive components which act at right-angles to the

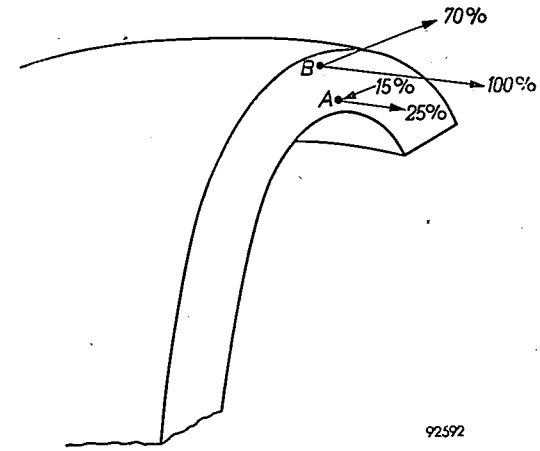


Fig. 3. Section of rim after cleating by old method; external diameter (before cleating) 4 mm. At A the radial and tangential deformations are about 15% and 25% respectively; at B about 70% and 100%.

tensile components. (An analogous state is produced in deformation processes such as cold-rolling and riveting.)

The first stage in the approach to the problem was, therefore, to design the cleat and the tool in such a way as to result in a smaller tangential deformation and, at the same time, greater pressure in radial directions. This implies keeping the material of the rim more bunched together. With that object in view the profile of the tool was modified, and experiments showed that the profile indicated in *fig. 4* produced the best results. It was also found better to make the rim of the cleat slightly thicker than before. Apart from questions of design, various kinds of materials and different lubricants were included in the investigations; the optimum duration of the cleating process was determined and, for the spinning method, the optimum speed of revolution. All experiments were carried out in accordance with a statistically planned schedule [1]).

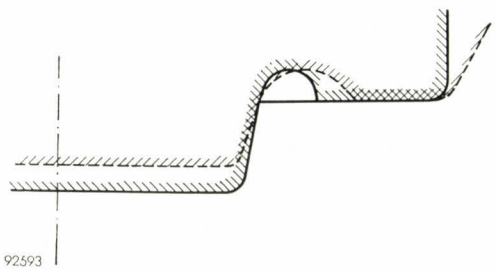


Fig. 4. Old profile (dashed) and new profile (full line) of cleating tool.

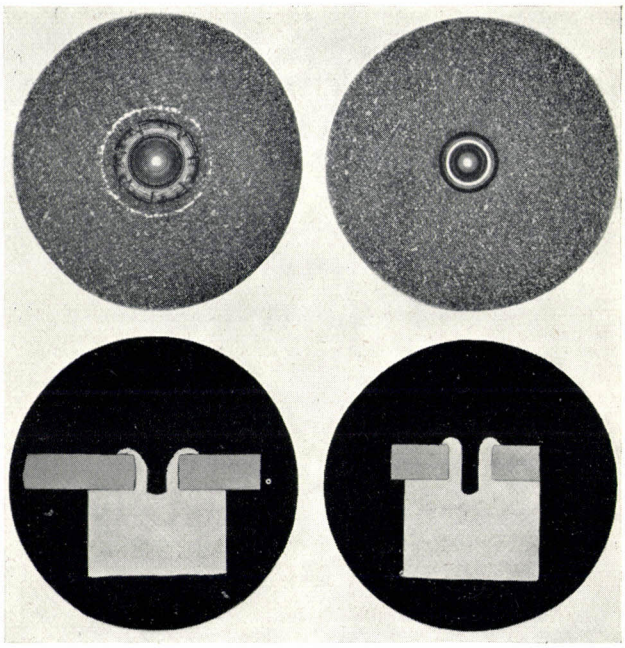


Fig. 5. Cleated joints seen from above and in section. Material: brass. Magnification: 3×. *Left:* example of joint made by old method and which frequently had to be rejected because of cracked rim. *Right:* reproducible cleated joint made by improved method.

The result was an appreciable improvement in the appearance of the cleated joints — as demonstrated in *fig. 5* — and in their mechanical strength, achieved without the need for the various pretreatments referred to above.

H. E. DEELMAN.

[1]) H. E. Deelman, Sigma **3**, 26-29, 1957 (No. 2).

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.

2494: P. Jongenburger: Energy of formation of vacancies in copper and gold (Phys. Rev. **106**, 66-69, 1957, No. 1).

From the anomalous rise of the thermal expansion near the melting point, a value of about 0.7 eV for the energy of formation of vacancies in copper and gold is found.

2495: H. A. Klasens: Elektroluminescentie (Ingenieur **69**, E.77-E.80, 1957, No. 23). (Electroluminescence; in Dutch.)

The construction of an electroluminescent panel is described. The properties of electroluminescent phosphors and the mechanism of electroluminescence are briefly discussed. Possible applications of electroluminescence are mentioned.

2496: J. L. Meijering: Calculation of the nickel-chromium-copper phase diagram from binary data (Acta metallurgica **5**, 257-264, 1957, No. 5).

In the solid state the Ni-Cr-Cu system shows a three-phase equilibrium between a Cr-rich b.c.c. phase and two f.c.c. phases rich in Ni and in Cu respectively. This rather unexpected feature (Ni and Cu being completely miscible) is examined thermodynamically, using the regular approximation. The most probable values of the interaction

parameters are obtained from data on the binary boundary systems. The resulting ternary diagram is indeed of the type found experimentally.

2497: G. Diemer, H. A. Klasens and P. Zalm: Voltage dependence of electroluminescence (J. Electrochem. Soc. **104**, 130 C, 1957, No. 6).

Re-examination of the results of Lehmann shows that they fit an exp $(-b/V^{\frac{1}{2}})$ brightness law over a larger range than suggested by this observer. This is in agreement with measurements by the present authors.

2498: A. van Weel: The design of phase-linear intermediate-frequency amplifiers (J. Brit. Instn. Rad. Engrs. **17**, 275-286, 1957, No. 5).

Data are given for the design of i.f. amplifiers with flat group-delay characteristics for combinations of from one to six tuned circuits. Both staggered-tuned single circuits and combinations of band filters are considered. Figures of merit are calculated and compared with the figures for networks based on flat-amplitude design.

2499: K. Compaan and Y. Haven: Correlation effects in diffusion problems in solids (Trabajos III reunión internacional sobre reactividad de los sólidos, Madrid, April 1956, section I, pp. 255-269; published 1957).

The correlation, which in many cases exists between the consecutive jumps of a radioactive tracer diffusing into solids, affects in certain types of diffusion mechanisms in ionic solids the diffusion constant of tracers, but not the contribution of these ions to conductivity. As a consequence, the relation of Einstein contains a correlation factor f which depends on the crystal structure as well as on the mechanism of ionic transport. It will be shown how an accurate test of the relation of Einstein can give information on the mechanism of ionic transport through solids. The following examples are discussed: diffusion of Na^+ through NaCl, Ag^+ through silver halide, Li through germanium and sodium through glass.

2500: G. H. Jonker: Compounds in the system Li_2O-TiO_2 and their stability (Trabajos III reunión internacional sobre reactividad de los sólidos, Madrid, April 1956, section I, pp. 413-421; published 1957).

In the oxide system Li_2O-TiO_2 three compounds are found: Li_2TiO_3, $Li_4Ti_5O_{12}$ and $Li_2Ti_3O_7$. The first compound can be prepared in an unstable unordered form with NaCl structure but prolonged heating leads to a stable orthorhombic pseudohexagonal super-structure. For the second compound, which has a spinel structure, it is established that the composition is not $Li_4Ti_7O_{16}$ as has been mentioned in the literature but a more normal one with a metal-to-oxygen ratio of 3 : 4. The third compound has an unknown crystal structure and can be obtained e.g. from the former one by evaporation of Li_2O at high temperature.

2501: A. L. Stuijts and H. B. Haanstra: Grain growth in a ceramic material, studied by means of electron microscopy (Trabajos III reunión internacional sobre reactividad de los sólidos, Madrid, April 1956, section I, pp. 671-701; published 1957).

Description of a replica technique whereby it is possible to make electron-microscope studies of changes in small pre-specified regions of a surface. Observations can be repeated on exactly the same region without damaging the surface and allowing treatment of the surface in between each observation. A series of electron micrographs are shown of a ceramic material ($BaFe_{12}O_{19}$) whose surface shows large micro-relief and which is somewhat porous. The photographs show that it is possible to follow changes in a group of crystallites after sintering for various times and at various temperatures.

2502: G. H. Jonker, H. P. J. Wijn and P. B. Braun: A new class of oxidic ferromagnetic materials with hexagonal crystal structures (Proc. Instn. Electr. Engrs. **104 B**, suppl. No. 5, 249-254, 1957).

Several new groups of ferromagnetic oxides have been found. They have rather complicated chemical compositions and can be obtained as reaction products of Fe_2O_3, BaO and MeO in various proportions, where Me denotes, in general, an element of the first transition group, or either Mg or Zn. The crystal structures of these new compounds can be described in the hexagonal system and are related to the magnetoplumbite structure. The crystal anisotropy of many of these hexagonal compounds is such that, depending on the chemical composition, the spontaneous magnetization in each crystal is oriented either parallel or perpendicular to the hexagonal axis. In the latter case the magnetization can fairly easily be rotated in the basal plane. Polycrystalline sintered samples of such compounds show an initial permeability of about 10; the permeability remains substantially constant up to about 1000 Mc/s, so that these new materials extend the frequency range where ferromagnetic oxides can be advantageously used as core materials to

frequencies much higher than is now possible with ferrites having cubic crystal structures. An earlier paper on these materials has been published in Philips tech. Rev. **18**, 145, 1956/57 (No. 6).

2503: E. W. Gorter: Saturation magnetization of some ferrimagnetic oxides with hexagonal crystal structures (Proc. Instn. Electr. Engrs. **104 B**, suppl. No. 5, 255-260, 1957).

Anderson's treatment of super-exchange yields a qualitative rule for the dependence of interaction on angle Me-O-Me. Using this rule, and assuming a longer distance to lead to a weaker interaction, low-temperature saturation magnetizations for ferrimagnetic oxides with known crystal structures can often be predicted. The results obtained for some of the materials described elsewhere (No. **2502**, these Abstracts), i.e. $Ba_2Me^{II}_2Fe_{12}O_{22}$ (**Y**), $Ba_3Me^{II}_2Fe_{24}O_{41}$ (**Z**), and $BaMe^{II}_2Fe_{16}O_{27}$ (**W**) (in which Me^{II} may be Mn, Fe, Co, Ni, Mg, or Zn), together with those for $BaFe_{12}O_{19}$ (**M**) and $KFe_{11}O_{17}$, with related crystal structures, are compared with the results of saturation-magnetization measurements down to 20 °K. It is shown that very small differences in crystal structure, such as exist between $KFe_{11}O_{17}$ and **M**, and between **M** and **Y**, can result in drastic differences in spin orientations of neighbouring ions. In particular for **Y**, the experimental data yield some insight into the distribution of the divalent Me ions among the various lattice sites. The crystal structure of **Z** is composed of alternating blocks of **M** and **Y**, that of **W** is composed of alternating blocks of **M** and **S** (spinel). The saturation magnetizations are obtained by adding those of the composing blocks.

2504: H. Koelmans and A. P. M. Cox: Luminescence of modified tin-activated strontium orthophosphate (J. Electrochem. Soc. **104**, 442-445, 1957, No. 7).

The introduction of Al, Zn Mg, Ca, or Cd in $Sr_3(PO_4)_2$ produces a new crystal phase which is probably isomorphous with $\beta Ca_3(PO_4)_2$. Activated with tin this Sr-orthophosphate gives a strong luminescence with a maximum at about 6300 Å under excitation with 2537 Å.

2505: H. Koopman, J. H. Uhlenbroek and J. Daams: New s-triazine herbicides (Nature **180**, 147-148, July 20, 1957).

Note on preparation of 2-alkoxy and 2-alkylthio 4,6-dichloro-s-triazines. It is found that the chlorine atom can be replaced by an alkylamino, alkyloxy or alkylthio group. Certain of these compounds show strong phytotoxic activity. Some preliminary tests are described.

2506: G. H. B. Teunissen, O. Bosgra and J. H. G. Wilson: Praktijkervaringen met levend, avirulent hondeziektevirusvaccin (T. Diergeneesk. **82**, 493-506, 1957, No. 14). (Practical experiences with live avirulent canine-distemper virus vaccine; in Dutch.)

Two series of tests on the vaccination of dogs against distemper and hard-pad disease. In all, 584 dogs from surroundings where canine-distemper is of frequent occurrence, were vaccinated with live avirulent canine-distemper virus vaccine. No data were obtainable with regard to 17 of these dogs. The subsequent history of the remaining 567 was traced for a considerable period, some up to two years after vaccination. Results show that puppies can, without danger, be immunized with live avirulent canine-distemper virus vaccine against both distemper and hard-pad disease virus infections.

2507: J. A. Kok: Doorslagverschijnselen in vloeibare isolatoren (Ingenieur **69**, E.87-E.92, 1957, No. 27). (Electric breakdown in liquid insulators; in Dutch.)

Polarizable material drifts towards a place of maximum stress by the action of transverse gradient $-E^2$ — forces and forms a bridge, if $r^3E^2 > \frac{1}{4}kT$. The critical diameter of the particles at $T =$ 300 °K is for $E = 22$ kV/mm: $2r = 50$ Å and for $E = 94$ kV/mm: $2r = 20$ Å. The breakdown strength of Shell Diala C oil was increased considerably by purification in an electrostatic Cottrell filter.

2508: J. P. L. Bots: Preparation of pure lactulose (Rec. Trav. chim. Pays-Bas **76**, 515-518, 1957, No. 6).

A simple method is described for the preparation of lactulose octa-acetate by acetylation of the reaction product obtained by treating lactose with lime. Pure lactulose was prepared by de-acetylation of lactulose octa-acetate with 0.1 *N*-HCl at 71 °C.

2509: J. S. C. Wessels: Studies on photosynthetic phosphorylation, I. Photosynthetic phosphorylation under anaerobic conditions (Biochim. biophys. Acta **25**, 97-100, 1957, No. 1).

The photochemical esterification of inorganic phosphate into adenosine triphosphate by chloroplasts was investigated under anaerobic conditions in the presence of either vitamin K_3 or flavin mononucleotide. Evidence is presented in support of the conclusion that vitamin K_3 and flavin mononucleotide are involved in separate pathways for photosynthetic phosphorylation. A tentative scheme for the generation of adenosine triphosphate in chloroplasts, consistent with this conclusion, is given.

2510: J. M. Stevels: Quelques remarques sur la résistivité du verre en courant continu (Silicates industriels **22**, 325-335, 1957, No. 6).

An attempt is made to discuss the well-known equation for the specific resistance ϱ of glasses as a function of the temperature T: $\log \varrho = A + B/T$, by a model of potential wells separated by potential barriers. Formulae are given for the constants A and B, for silicate glasses containing only one kind of network-modifying ions, relating these constants with quantities such as jump distance, height of the potential barriers, concentration and vibration frequency of the network-modifying ions. The limits of the validity of the above equations in terms of the temperature and the electric field applied is also discussed. Finally the relation between A and B and the coherence of the Si-O network is discussed in an extensive way. It is found that the height of the potential barriers (activation energies of the moving network-modifying ions) is not a smooth function of the degree of cohesion of the network, but shows bends. The physical background of these bends is discussed.

2511: M. J. Sparnaay: Attractive forces between flat plates (Nature **180**, 334-335, Aug. 17, 1957).

Note concerning measurements using metal plates instead of the glass or quartz plates used previously. The results show that Casimir's d^{-4} relation is not contradicted.

R 328: G. Braun: Mathematical treatment of a typical zone-melting process (Philips Res. Rep. **12**, 385-414, 1957, No. 5).

This paper deals with the mathematical theory of a typical example of a zone-melting process as described by Van den Boomgaard in a preceding Report (No. **R 275**, these Abstracts). The process in question aims at the realization of a most homogeneous distribution of a solute (which may be volatile) in an ingot. This is effected by a large number of identical steps and amounts to an extension of the procedure proposed by Pfann and Reiss. The intermediate distributions tending to the limiting one are considered in detail, as well as the number of steps necessary to arrive at such a distribution. The mathematics determine the asymptotic behaviour of the solution of a system of differential-difference equations by means of a generating function.

R 329: G. Brouwer: Graphical analysis of the temperature dependence of the electronic population in semiconductors (Philips Res. Rep. **12**, 415-422, 1957, No. 5).

The distribution of electrons and holes in semiconductors and its variation with temperature may be derived by means of a graphical approximation presented in this paper. The method is based directly on the electronic reaction equations in the equilibrium.

R 330: G. D. Rieck: The texture of drawn tungsten wires (Philips Res. Rep. **12**, 423-431, 1957, No. 5).

According to a paper by Jeffries, which although old, is still referred to in textbooks, drawn tungsten wire has a texture with the [110] axis of the grains in the wire direction and their (001) planes parallel to the surface of the wire. It is proved, however, with the aid of X-ray diffraction measurements, that the tungsten wire has the generally accepted normal fibre texture with a random orientation of the (001) planes around the [110] fibre axis.

R 331: J. S. van Wieringen and A. Kats: Paramagnetic resonance and optical investigation of silicate glasses and fused silica, coloured by X-rays (Philips Res. Rep. **12**, 432-454, 1957, No. 5).

Silicate glasses containing alkali oxides (or mixtures of alkali and alkaline earth oxides) are irradiated with X-rays both at room and liquid-nitrogen temperature. Colour centres are produced whose optical and paramagnetic resonance properties have been studied. All silicate glasses show two paramagnetic resonance peaks due to two different colour centres containing electrons and holes, respectively. From the line widths at 9500 and 24000 Mc/s it follows that both centres have a spin $\frac{1}{2}$ and are in local electric fields of very low symmetry. Comparison with the optical measurements shows that the electron centre is likely to be connected with the optical absorption band between 4150 Å and 4900 Å and the hole centre with the band at 3100 Å. Similar measurements are made on fused silica irradiated with X-rays at room temperature. Again two paramagnetic resonance peaks are found, ascribed to electrons and holes, respectively. In the optical spectrum three absorption bands are developed. The behaviour of fused silica depends strongly on its water content. After irradiation the samples containing water show only the resonance peak of the electron centre and the optical absorption band at 2150 Å; the hole peak and the other two bands are absent or extremely weak.

VOLUME 19, 1957/58, No. 10 pp. 273-300 Published 30th April 1958

Philips Technical Review

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

MONITORING, CONTROL AND SAFETY EQUIPMENT FOR A NUCLEAR REACTOR OF THE SWIMMING-POOL TYPE

II. FURTHER DESCRIPTION OF CERTAIN COMPONENT UNITS

by F. E. L. ten HAAF, G. KLEIN and F. J. SCHIJFF. 621.316.7:621.039.4

The design of the auxiliary electronic equipment for a nuclear reactor involves problems which often differ appreciably from those commonly encountered in electronic engineering, and which therefore require a new approach. Some of these problems, which are dealt with below, concern the design of 1) a circuit giving a logarithmic relation between two quantities over a range of many decades, 2) a gamma indicator which gives a direct reading of the dose rate and which can also stop the reactor if necessary, and 3) a safety system which provides complete protection in the event of defects occurring in the reactor.

For the purposes of monitoring, control and safety a nuclear reactor requires elaborate equipment, which is largely electronic. In Part I of this article[1]) a general description was given of the equipment provided for the research reactor shown in operation last year at the exhibition "Het Atoom" at Amsterdam, and which is shortly to be installed at the Technische Hogeschool, Delft.

We shall deal here in more detail with three units of this equipment: 1) the logarithmic measuring channel for high neutron flux, 2) the gamma indicator, and 3) the safety amplifiers.

Logarithmic measuring channel for high neutron flux

The logarithmic neutron-flux and period meter, type PW X081 (*fig. 1*), which is connected to an ionization chamber in the reactor, is divided, according to its functions, into

a) a logarithmic amplifier,
b) a period amplifier,
c) a relay stage, and
d) various calibration devices.

[1]) M. van Tol, Monitoring, control and safety equipment for a nuclear reactor of the swimming-pool type, I. General description, Philips tech. Rev. **19**, 245-257, 1957/58 (No. 9).

The logarithmic amplifier must deliver an output voltage which is proportional to the logarithm of the current in the ionization chamber, and therefore to the logarithm of the neutron flux in the reactor. This output voltage causes the deflection of a meter which is provided with a logarithmic scale (to base 10). The scale, calibrated in percentages of the maximum permissible neutron flux, runs from 10^{-4} % to 300 %, thus covering $6\frac{1}{2}$ decades. Provision is made for the connection of a recording instrument.

In the period amplifier the output voltage of the logarithmic amplifier is differentiated and the result amplified. This produces a voltage which is inversely proportional to the "period" (time constant) T of the nuclear reaction.

Since the neutron flux N varies exponentially with the time t, i.e.

$$N = N_0 \, e^{t/T},$$

the logarithmic amplifier delivers a voltage proportional to $\ln N$, and therefore proportional to (const. $+ t/T$); thus by differentiation with respect to time t a voltage is obtained which is proportional to $1/T$.

The period T can be read directly from a meter having a reciprocal scale and calibrated in seconds. The scale runs from —30 seconds via infinity to +3

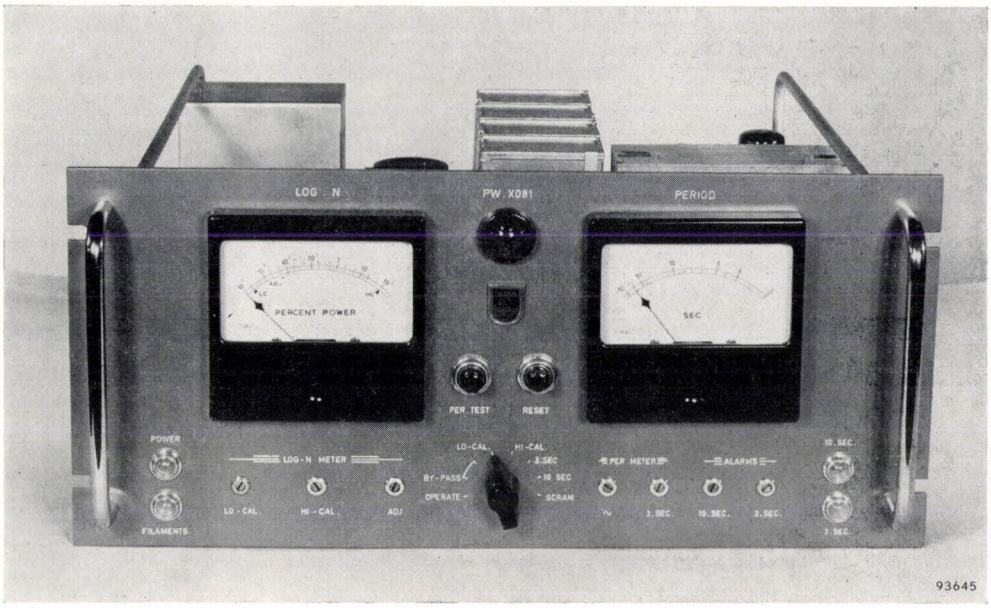


Fig. 1. Unit PW X081 of the logarithmic measuring channel for high neutron flux. Left, the neutron-flux meter with logarithmic scale from $10^{-4}\%$ to 300%; right, the period meter with scale from −30 via ∞ to 3 seconds. Below the meters, the correction resistors, adjustable with a screwdriver; between them a switch with eight positions (one operating position and seven calibration and test positions).

seconds (see fig. 10 in Part I). Here, too, a recording instrument can be connected.

The output voltage is fed to a relay stage as well as to the meters. This stage consists of two relays which come into operation as soon as the reactor period drops below 10 and 3 seconds respectively. The relay contacts ensure that the necessary warning signals are given and the appropriate action taken in the control system: when the period drops to 10 seconds the control rod and the safety rods descend deeper into the reactor core. The effect of this is to reduce the neutron multiplication factor k. If, in spite of this, the period goes on dropping and reaches 3 seconds, the safety rods drop right into the core and thereby stop the chain reaction.

By means of a switch and a push-button, certain standard signals can be supplied to the input to check the reading on both meters. Discrepancies can be corrected by adjusting a number of potentiometers.

The logarithmic amplifier

A familiar circuit element with logarithmic properties is the thermionic diode. Over a certain portion of its characteristic — the retarding-field region — the following relation exists between the current I and the voltage V:

$$I = I_0 \exp \frac{eV}{k\Theta}, \quad \ldots\ldots \quad (1)$$

where I_0 is a characteristic quantity for the cathode, e the electronic charge, k Boltzmann's constant and Θ the absolute temperature of the cathode. Equation (1) is derived from the fact that the electrons have a Maxwellian velocity distribution. The lower limit of the current range in which (1) is applicable is determined by the quality of the vacuum and of the insulation, and the upper limit is determined by the occurrence of space charge.

If (1) be written

$$V = \frac{k\Theta}{e} \ln I - \frac{k\Theta}{e} \ln I_0, \quad \ldots\ldots \quad (2)$$

a logarithmic indication of the current I is found by measuring the voltage V over the diode. In practice the change in V amounts to approximately 0.2 V when I changes by a factor of 10.

However, a fixed relation between V and I exists only if the cathode temperature Θ is constant. The second term on the right-hand side of (2) is especially sensitive to changes in Θ, since I_0 also increases with Θ. The filament voltage of the diode must therefore be very well stabilized.

This critical sensitivity to variations in the filament voltage can be circumvented by using a triode instead of the diode. With the triode, in the lower region of its grid voltage/grid current characteristic, an analogous exponential relation exists, as in the retarding-field region of the diode. Further, we may write for a triode, quite generally,

$$\delta I_a = S\,\delta V_g + \frac{S}{\mu}\,\delta V_a, \quad \ldots\ldots \quad (3)$$

where I_a is the anode current, V_g the grid voltage, V_a the anode voltage, S the slope(transconductance) and μ the amplification factor. When the anode voltage is constant, (3) reduces to

$$\delta I_a = S\,\delta V_g.$$

A logarithmic relationship also exists, therefore, between I_a and I_g, provided at least that S is constant in the region concerned. In practice this region covers about three decades.

To ascertain the effect of a change in the cathode temperature on the functioning of the circuit, let the filament voltage be raised slightly and consider what happens to the anode current, the grid current being kept constant. Analogous to the case of the diode, both I_g and I_a tend to rise owing to the increase in the cathode temperature. However, since I_g is kept constant, the grid voltage must go more negative, and this in turn causes I_a to decrease. Thus, I_a is subjected to two opposing influences, the one tending to increase it and the other to decrease it. Closer investigation shows that these influences largely compensate each other, so that filament voltage variations cause no appreciable change in the anode current. In this case, then, a rough stabilization of the filament is sufficient.

Another method of obtaining a logarithmic characteristic with a triode is to measure the anode voltage at constant anode current as a function of the grid current. In that case, $\delta I_a = 0$, so that (3) changes to

$$\delta V_a = -\mu\,\delta V_g.$$

Provided μ is constant (this region being larger than that in which S is constant), proportionality thus also exists between δV_a and $\ln(\delta I_g)$. In a manner analogous to that described above, changes in filament voltage are again compensated. A resistance R_a, large with respect to the internal resistance μ/S of the triode, connected to a fixed positive potential (*fig. 2*), as is usual in voltage amplifiers, ensures a practically constant anode current I_a.

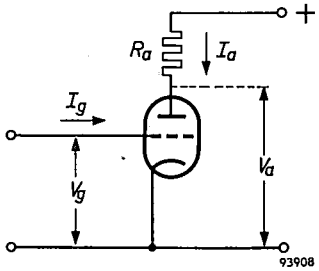


Fig. 2. A logarithmic relation exists between the anode voltage V_a and the grid current I_g of a triode operating in the lower region of its characteristic, at constant anode current I_a. The anode current is here kept approximately constant by the high resistance R_a in the anode circuit.

A tube which is found to be very efficient in the latter circuit is the type E 80 F pentode, operated as a triode. The anode voltage of this tube changes by about 4 V per decade change of grid current. It is not difficult to find a tube of this type whose V_a-I_g characteristic is logarithmic over a grid-current range of $6\frac{1}{2}$ decades.

The period amplifier

As stated, the output voltage of the logarithmic amplifier must be differentiated in order to obtain a voltage which is proportional to $1/T$ and therefore a measure of the period T.

A familiar differentiating network, consisting of a capacitance C and a resistance R, is shown in *fig. 3a*. It is not, however, an ideal differentiator, for

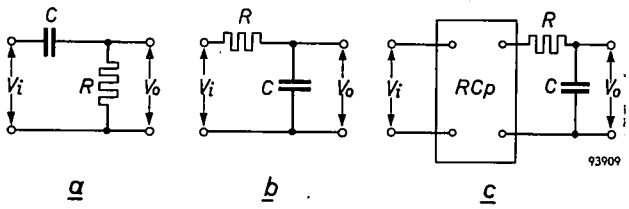


Fig. 3. *a*) Non-ideal differentiator. *b*) Non-ideal integrator. *c*) The non-ideal differentiator (*a*) is equivalent to the ideal differentiator RCp and the non-ideal integrator (*b*) in cascade.

a discontinuous change in the derivative of the input voltage V_i (*fig. 4a*) does not cause a jump in the output voltage, as in fig. 4*b*, but a more gradual change moving exponentially to its asymptotic value (fig. 4*c*), the time constant being RC. For our purpose this is undesirable, since the period meter is required to follow, with the least possible delay, every change in the derivative of the output voltage from the logarithmic amplifier. The smaller the time constant RC, the shorter the delay. However, a reduction of RC entails a drop in the output voltage V_o (it is obvious that $V_o = 0$ in the extreme case $R = 0$ or $C = 0$).

The following considerations will help to make this clear. For the differentiating network in fig. 3*a* the differential equation relating the output voltage V_o and the input voltage V_i is

$$V_o + RC\frac{dV_o}{dt} = RC\frac{dV_i}{dt},$$

or, replacing d/dt by the Heaviside operator p:

$$V_o = \frac{1}{1+RCp}\,RCpV_i, \quad \ldots\ldots \quad (4)$$

whereas for an ideal differentiator

$$V_o \propto pV_i.$$

For the analogous integrating network (fig. 3*b*):

$$V_o + RCpV_o = V_i,$$

hence

$$V_o = \frac{1}{1+RCp}\,V_i. \quad \ldots\ldots \quad (5)$$

Comparing (4) with (5), we see that the only difference consists in the factor RC_p, which occurs in (4) but not in (5). This means that (4) corresponds to two networks in cascade, the first being an ideal differentiator RC_p and the second the integrator from fig. 3*b*. In other words, fig. 3*c* is equivalent to fig. 3*a*.

It is clear from fig. 3*c* that the smaller R and C become, the closer is the approximation to an ideal differentiator, but it follows from (4) that making RC smaller causes the output voltage V_o to decrease.

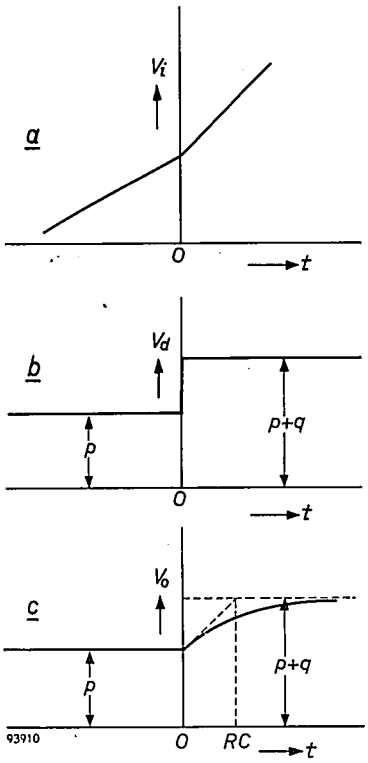


Fig. 4. *a*) Voltage V_i linear with time t, but with a discontinuity in the derivative at $t = 0$:

$$t < 0 \quad . \quad . \quad . \quad V_i = pt,$$
$$t > 0 \quad . \quad . \quad . \quad V_i = (p + q)t.$$

b) Output voltage V_d of an ideal differentiator whose input voltage has the form shown in (*a*).
c) Output voltage V_o of a non-ideal differentiator as in fig. 3*a*, with input voltage of the form (*a*). The output voltage for $t > 0$ is given by

$$V_o = RC\left[p + q\left\{1 - \exp(-t/RC)\right\}\right].$$

When the period T is 10 seconds, the logarithmic amplifier delivers a voltage which increases by about 0.2 V per second. If we take for RC a value of, say, 0.1 second (the time constant should certainly not be much larger), then, where $T = 10$ seconds, the output voltage of the differentiator is only 20 mV. This is too low for operating the relay stage, for which reason an amplifier is interposed. This consists of two stages (*fig. 5*), the second being a triode cathode follower.

If this amplifier were simply to be placed behind the differentiator (*fig. 6a*), its output voltage — and consequently the reading on the period meter —

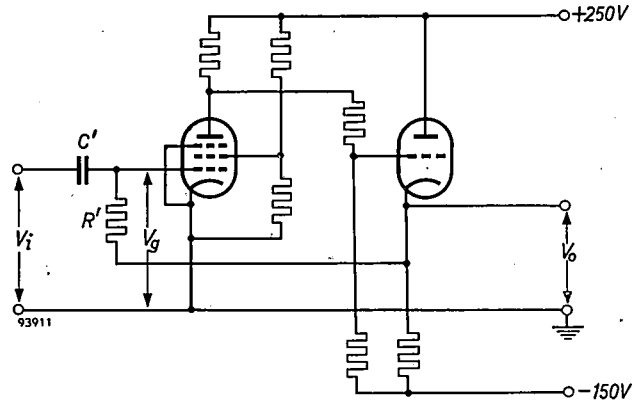


Fig. 5. Differentiator (C' and R') with amplifier. R' is connected to the output (triode cathode follower).

would be subject to changes in the amplification. There is a simple way, however, of making the output voltage practically independent of the amplification, namely by connecting the resistor in the differentiator to the output of the amplifier instead of to earth (fig. 6*b*, cf. also fig. 5) and by fulfilling the following conditions:

1) the elements R' and C' of the differentiator should be chosen such that $R'C' = A_0RC$, where A_0 is the nominal value of the amplification and RC the original time constant;
2) the amplification A must be much larger than unity.

Fig. 7b shows the step-function response of a circuit which satisfies these conditions. It can be seen that the output voltage, irrespective of A, approaches the same final value (which is not the case with the circuit in fig. 6*a*; see fig. 7*a*). The way in which the final value is approached now depends on A, but this simply means that the period meter

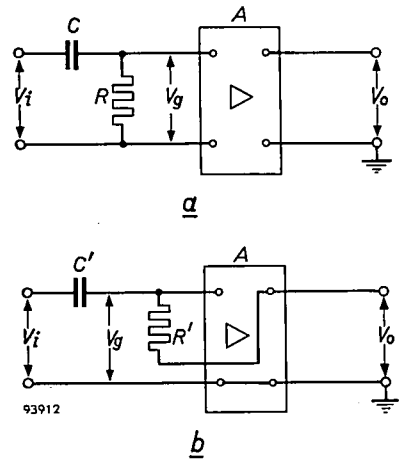


Fig. 6. *a*) Differentiator (C and R) followed by an amplifier with amplification A. The output voltage is proportional to A. *b*) Differentiator (C' and R') and amplifier, with R' connected to the amplifier output. Under certain conditions the output voltage is practically independent of A.

reaches its final value rather more quickly or slowly, depending upon A.

The following calculation will serve to explain this. In fig. 6*a* the differentiator has V_i as input voltage and $V_g = -V_o/A$ as output voltage (with this choice of signs, A is a positive number). According to (4),

$$V_o = -A \frac{1}{1 + RCp} RCp V_i , \quad \ldots\ldots \quad (6)$$

which shows the proportional relationship between V_o and A.

In fig. 6*b*, on the other hand, $V_i - V_o$ is the input voltage and $V_g - V_o$ the output voltage of the differentiator, the time constant being $R'C'$. With $V_g = -V_o/A$ and after some manipulation, equation (4) now gives:

$$V_o = -\frac{\frac{A}{A+1}}{1 + \frac{R'C'p}{A+1}} R'C'p V_i . \quad \ldots\ldots \quad (7)$$

Putting $R'C' = A_0RC$, we may write (7) as

$$V_o = -\frac{A}{A+1} A_0 \frac{1}{1 + \frac{A_0}{A+1} RCp} RCp V_i .$$

If, moreover, the amplification is made so large that $A/(A+1) \approx 1$, then

$$V_o \approx -A_0 \frac{1}{1 + \frac{A_0}{A} RCp} RCp V_i . \quad \ldots\ldots \quad (8)$$

Eq. (8) is of the same form as eq. (6), with the important difference, however, that the constant factor A_0 has taken the place of the fluctuating factor A. Furthermore, RC in the denominator has been replaced by the time constant A_0RC/A, which is not independent of A. As remarked above, however, this is of no practical importance.

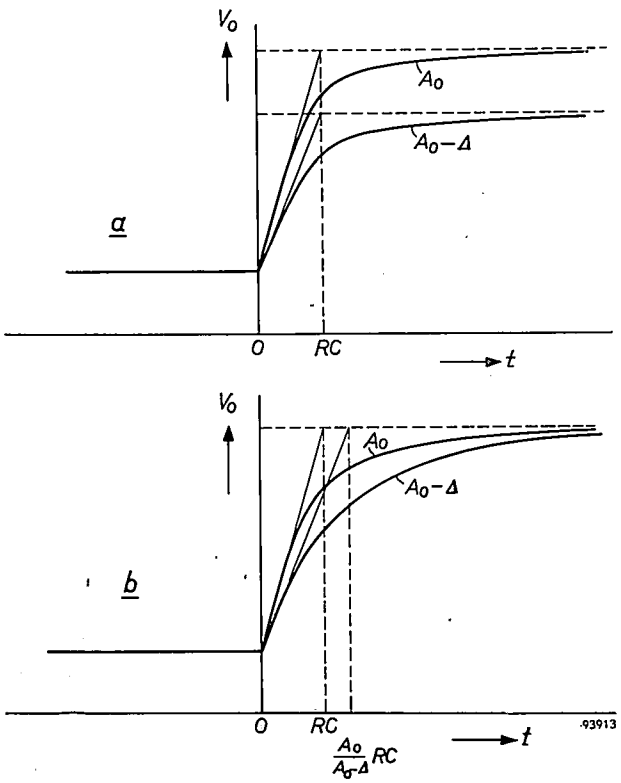


Fig. 7. *a*) Step-function response of the circuit in fig. 6*a*, for two values of the amplification, A_0 and $A_0 - \Delta$. The end values differ, the time constant in both cases is RC.
b) The same for the circuit in fig. 6*b*, with $R'C' = A_0RC$ and $A \gg 1$. Now the time constants differ, but the end values are equal.

The relay stage

The relay stage (*fig. 8*) contains two double triodes operating as so-called Schmitt triggers, *I-II* and *III-IV*. These are so biased that in the absence of a signal from the period amplifier (i.e. with the reactor in the stationary or shut-down state, period $T = \infty$) only *II* and *IV* pass current. In this state the relays A_1 and A_2 are energized; the positions of contacts a_{11}, a_{12}, a_{21}, a_{22} and a_{23} are then the reverse of those drawn in fig. 8.

When the neutron flux increases, T has a finite positive value and the period amplifier delivers a positive output voltage, which is higher the shorter the period. By means of potentiometers P_1 and P_2 respectively, each of the circuits can be so adjusted as to make the loop gain greater than 1 at a given value of T. For the pair *I-II* this point is fixed at $T = 10$ seconds. When this is reached the circuit triggers over into the state in which *I* is conducting and *II* is cut off. The relay A_1 cuts out and contact a_{12} opens. As a result, a warning signal is given on the control desk, and via the intermediary of the servo-mechanism the control rod and the safety rods in the reactor start descending to counteract any further increase in the neutron flux.

If, nevertheless, the period goes on becoming shorter and falls to 3 seconds, *III* becomes conducting and *IV* is cut off. The effect of this is to open contact a_{23}, which interrupts the current to the magnets. As a result, the safety rods drop into the core and the reactor is stopped. At the same time, contact a_{22} opens and a warning signal is given. Moreover, contact a_{21} breaks the anode lead to *IV*, so that relay A_2 can only be actuated again when button *D* is depressed.

The calibration devices

By means of an eight-position switch (fig. 1) certain calibrations and checks can be carried out. As soon as the switch is turned from the measuring (normal) position, a red pilot lamp lights up on the front panel; at the same time a contact is closed which switches on a signal elsewhere, in this case on the control desk. To make calibration possible while the reactor is in operation, in positions 2 to 7 the relay contacts a_{12}, a_{22} and a_{23} are short-circuited (if they were open they would switch on the safety amplifiers and start the servomotors). All calibration adjustments can be made at the front panel with a screwdriver.

The positions of the switch are the following:

1) Normal position (measuring position).
2) In this position everything functions normally, except that the relay contacts mentioned are short-circuited. This is necessary to prevent a false alarm from being sounded when switching back from the calibration positions to the measuring position.
3) As (2), but now the period meter is also short-circuited.
4) Here a known current (10^{-10} A) is fed to the logarithmic amplifier, which should produce a reading on the meter of 2×10^{-4} % of full power.

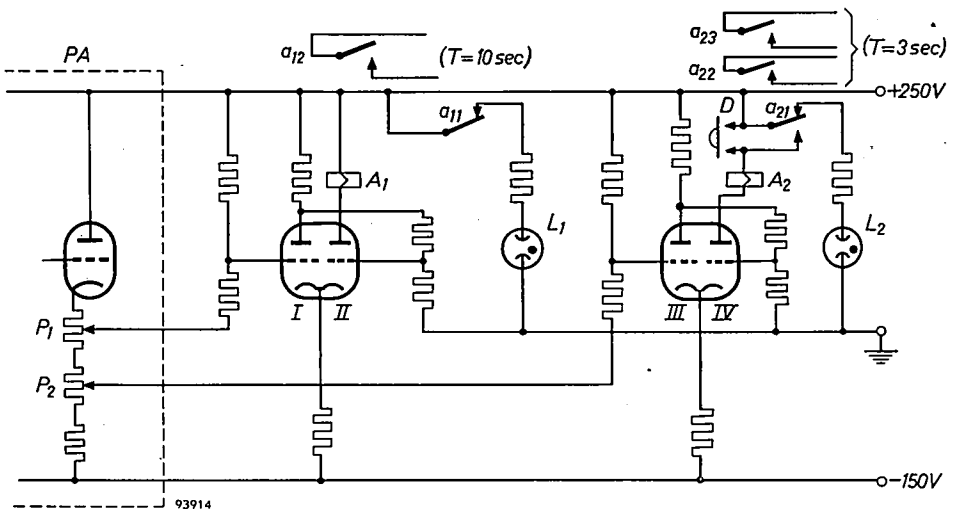


Fig. 8. Relay stage. When the period amplifier PA supplies no signal ($T = \infty$), the right halves (II, IV) of the two double triodes both pass current and the relays A_1 and A_2 are energized; the positions of contacts a_{11}, a_{12}, a_{21}, a_{22} and a_{23} are then the reverse of those shown here. If T drops to 10 seconds, I becomes conductive and II is cut off; a_{11} closes, a_{12} opens. If T drops to 3 seconds, III becomes conductive and IV is cut off; a_{21} is switched over, a_{22} and a_{23} are opened. The value of T at which these operations take place can be adjusted with potentiometers P_1 and P_2 in the period amplifier. L_1 and L_2 are neon lamps which indicate the tripping of the relays.

5) As (4), but the current is now 10^{-5} A, and the calibration point lies at 20% of the full power.
6) When the button "Period test" is depressed, a voltage linear with time appears across the input of the period amplifier, the rate of increase corresponding to $T = 3$ seconds. The period meter should now indicate 3 seconds. At the same time the relay circuit III-IV triggers; this is indicated by the lighting-up of the neon bulb L_2 (fig. 8).
7) As (6), but now with a voltage increasing at a rate corresponding to $T = 10$ seconds. The neon bulb L_1 (fig. 8) lights up to indicate that the circuit I-II has triggered.
8) As (6), but now the relay contacts are no longer short-circuited. This is for checking whether the connections with the safety amplifiers and the servomotors are in order.

The gamma indicator

A nuclear reaction is accompanied by the generation of gamma rays. These rays are a hazard to the health of persons exposed to them. For this reason, and also to indicate the presence of leaks in the fuel elements, a constant check is kept on the intensity of the gamma radiation around the reactor; at the exhibition "Het Atoom" this was done at three positions. If the gamma radiation exceeds a certain intensity, the reactor is automatically shut down ("slow scram") and at the same time a warning signal is given.

Gamma rays, like X-rays, have the property of forming ions in the material through which they pass. The röntgen unit, used to express dosage of X-rays and gamma rays, is defined on the basis of this property. The röntgen is defined as follows (Report International Commission on Radiological Units, 1956): "One röntgen (1 r) is the exposure dose of X- or gamma radiation such that the associated corpuscular emission per 0.001293 gramme of air produces, in air, ions carrying one electrostatic unit of quantity of electricity of either sign". (0.001293 gramme is the mass of 1 cm³ of dry atmospheric air at 0 °C and 760 mm of mercury pressure.)

The International Commission for Radiological Protection has recommended that the dose rate (dose per unit time) received by persons exposed to radiation in their work should not exceed 0.3 r per week [2]).

[2]) See Philips tech. Rev. **16**, 136, 1954/55.

So far as is known, a normal person exposed to such a dose rate will suffer no harmful effect at any time in his life. In the case of the reactor at the Amsterdam exhibition it was assumed that the operating personnel could be exposed to the radiation for 30 hours a week maximum. It was therefore stipulated that at no accessible place should the dose rate exceed 10 milliröntgens per hour (10 mr/h). For comparison it may be noted that the dose rate of cosmic radiation at sea level is approximately 0.01 mr/h, i.e. about 1/1000 of the above stipulated limit.

The gamma detector

For detecting the gamma rays an ionization chamber is used, consisting of a thin-walled steel cylinder of 800 cm³ volume, filled with argon under 15 atm pressure [3]). The ions released by the gamma rays in the gas are removed by an electric field applied between a central electrode and the wall. The strength of the field ensures that almost no ions are lost by recombination (argon as a filler gas has the advantage over air of allowing the necessary field strength to be reached at a low voltage). The field strength is not so high, however, as to enable new ions to be formed by collisions with the gas atoms. The current obtained is thus dependent solely on the dose rate.

If the ionization chamber were filled with air at 1 atm and 0 °C, and if the volume were 1 cm³, the gamma rays at the maximum permissible dose rate of 10 mr/h would liberate per hour a charge of 0.01 e.s.u. $= 3.33\times10^{-12}$ coulomb. The maximum detector current would then be $3.33\times10^{-12}/60^2 \approx 10^{-15}$ A. In the actual detector this current is a great deal higher, namely (ignoring, among other things, the wall effect) by a factor of 800 due to the larger volume, a factor of 15 due to the higher pressure, and a factor of about 1.25 due to the use of argon instead of air (which is thus an incidental advantage of argon over air). The detector current at the maximum permissible dose rate is therefore approximately:

$$800\times15\times1.25\times10^{-15} \approx 1.5\times10^{-11} \text{ A.}$$

At the position of the ionization chamber the radiation has already covered an appreciable distance through water and possibly through concrete, so that what is left is practically only hard radiation. The power of hard-radiation quanta to dislodge electrons from the gas atoms is proportional to the number of electrons per cm³, and therefore to the number of gas atoms per cm³ and to the atomic number of the gas. The atomic number of nitrogen is 7, that of oxygen 8 (average for air: 7.2), and that of argon 18. Since 1 cm³ argon at 15 atm contains just as many atoms as there are molecules in 1 cm³ air at 15 atm, and since this number of molecules is half as large as the number of N and O atoms, the gain by using argon instead of air is $18/(2\times7.2) = 1.25$. (This simplified calculation does not include all the factors of influence.)

The minimum detector current required to be measured corresponds to the dose rate of cosmic radiation (approx. 0.01 mr/h), viz. a current of the order of 10^{-14} A.

Measurement of the detector current

For measuring the gamma-detector current the tetrode electrometer tube 4066 is used. The control grid is connected only to the central electrode of the gamma detector. The tube thus operates with a grid current which is equal and opposed to the detector current (the insulation is so efficient that the maximum leak current is of the order of 10^{-15} A and may therefore be neglected).

At this very weak grid current (10^{-14}-10^{-11} A, entirely in the exponential region), we again find that a change in the grid voltage δV_g is proportional to the logarithm of the associated change in the grid current, δI_g. Since δV_g results in a proportional change, δI_a, in the anode current, δI_a is proportional to log (δI_g). The grid-current changes, which are equal to the detector-current changes and thus proportional to the dose rate, can be read on a logarithmic scale on a moving-coil meter in the anode circuit (the relatively high quiescent anode current is compensated in the meter by a simple circuit arrangement). A logarithmic scale has the advantages, mentioned earlier in this article, that it covers many decades (so that one measuring range suffices, reducing the risk of mistakes in reading) and that the relative sensitivity is equal over the whole scale.

The electrometer tetrode type 4066 is a subminiature tube whose second grid acts as control grid. The first grid serves as space-charge grid, for which purpose it receives a constant positive voltage of a few volts.

The tube is mounted inside the gamma detector. The great advantage of this is that the connection of the central electrode to the control grid does not have to be led through the steel wall of the detector, which means that the part of the circuit which must be highly insulated is situated in a dry and unchanging environment. Lead-ins are now necessary, of course, for the supply to the other electrodes, but no particularly high insulation resistance is needed for these.

[3]) This gamma detector was designed and built by K. van Duuren and co-workers, for the Philips Research Laboratories, working at the synchrocyclotron at Amsterdam.

Transistorized light-relay

Since the anode current of the electrometer tube is too weak to energize a relay directly, a light-relay device is used.

At the end of the scale of the anode-current meter, at 10 mr/h, an opening is made in the dial. Through this opening, light from an electric bulb falls on to a transistor, which then passes a current between emitter and collector powerful enough to actuate a relay. As soon as the light is intercepted by the needle, however, the transistor current drops practically to zero; the relay then cuts out, and as a result the reactor is stopped and an alarm device deflection at 10 μA. The meters are horizontal because if such sensitive instruments were to be mounted upright (with the axis of the moving coil horizontal), the needle might easily stick at the beginning or the end of the scale. To facilitate the reading, however, a further meter M_2 is mounted on the front panel, which can be switched to each of the four channels (if this meter should stick, the safety measures are not affected).

The circuit of one channel can be seen in *fig. 10*. The correction resistor R_1 serves for matching the deflection of meter M_1 to that of M_2 from which the reading is made.

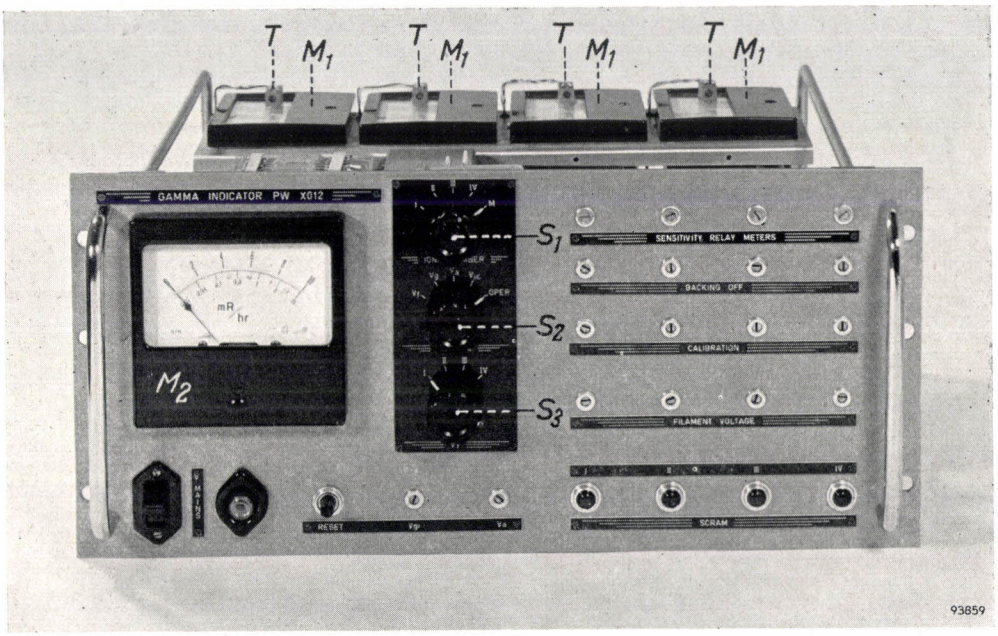


Fig. 9. Gamma-indicator unit PW X012. Above: the four light-relay meters M_1 on which are mounted transistors T. On the front panel: meter M_2 indicates the dose rate, or functions as a test voltmeter, depending on the position of switches S_1, S_2 and S_3.

set in operation. (In this arrangement, then, an amplifier, as would be needed with a normal photocell, is unnecessary.) To check whether the system is functioning correctly, a radioactive specimen can be placed near one of the ionization chambers.

In principle, every transistor is sensitive to light. In ordinary operation this phenomenon would be troublesome, which is why normal transistors are protected against light by a coating of lacquer. The OCP 71 transistors used here have no such coating, but are otherwise of the normal type.

Circuit of the gamma indicator

The panel of the gamma-indicator unit is shown in *fig. 9*. The circuit is equipped for four ionization chambers. The four light-relay meters M_1 can be seen with their transistors T. These meters give full

The relative sensitivity is 1 μA per factor 2, i.e. the current through M_1 increases by 1 μA when the dose rate is doubled. There should be full deflection (10 μA) at 10 mr/h. The range of the scale is therefore ten factors of 2, i.e. $2^{10} \approx 1000$. The zero point of the meter thus corresponds to 0.01 mr/h (cosmic radiation). The required relative sensitivity of 1 μA per factor 2 can be adjusted by regulating the voltage on the first grid (with series resistor R_3) and if necessary by varying the filament current (with resistor R_4).

After this adjustment, an absolute calibration is necessary, that is to say the anode current must be compensated in such a way that the meter will indicate the actual dose rate at the position of the detector. For this calibration a radioactive source

of known intensity is needed (the intensity may be very low, owing to the high sensitivity of the instrument). The compensation current can be varied with potentiometer P (fig. 10). Like the resistors R_1, R_3 and R_4, P can be adjusted with a screwdriver.

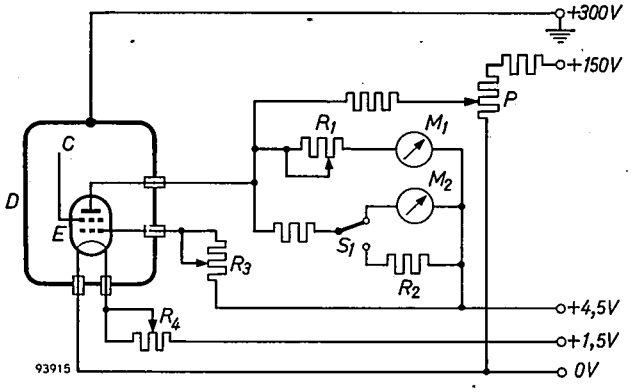


Fig. 10. Diagram of one channel of the gamma indicator. D is the ionization chamber functioning as gamma detector; mounted in it are the electrometer tetrode E (type 4066) and the central electrode C. M_1 light-relay meter. R_1 correction resistor. M_2 meter for reading the dose rate. S_1 switch used either for switching M_2 successively into the four channels or for measuring the supply voltages. R_2 equivalent resistor for M_2. P potentiometer for adjusting the compensation current. R_3 resistor for varying the voltage on the space-charge grid of the electrometer tube. R_4 filament resistor.

The meter M_2 can be switched into each of the four measuring channels by means of a five-position switch. An equivalent resistance R_2 takes the place of the meter in the channels in which the meter is not being used. The fifth position serves for checking the supply voltages on the same meter. These are stabilized such that mains voltage fluctuations up to 10% have no significant effect on the reading.

The steel detector cylinders are earthed for safety reasons. The other parts of the apparatus are thus at a potential of about —300 V with respect to earth.

Measurement with different radiation sources

Because of the use of argon instead of air in the ionization chambers, it might be feared that the ion yield would depend in a different way on the hardness of the radiation. Since, in our case, only hard radiation is present near the ionization chambers (the softer rays are mostly absorbed in the water and in the concrete), any such effect may be expected to be very small. To verify this, we carried out an experiment with two different radiation sources: cobalt (radiations of 1.1 and 1.3 MeV) and caesium (radiation of 0.66 MeV). At these energies argon was found to be practically equivalent to air.

After all specimens had been removed — leaving only the cosmic radiation — the meter reading was 0.01 mr/h. This demonstrates the highly effective insulation of the central electrode and the control grid.

The safety amplifiers

The safety system has two ionization chambers of its own. The current from each is fed to two "safety amplifiers", which function as relays. The system also comprises a square-wave generator (2000 c/s) and two rectifiers, one of which supplies the direct voltage (300 V) for the square-wave generator and for the ionization chambers, and the other the energizing current for the electromagnets from which the safety rods are suspended. All these parts are mounted in the safety unit type PW X086 [4]). If the current from one of the ionization chambers is too high — owing to the neutron flux being too high — one of the safety amplifiers interrupts the magnet currents, causing the safety rods to fall and thereby stopping the reactor.

The ionization chambers (of American manufacture) employed in the safety channels are of the parallel circular plate type ("PCP"). The electrodes consist of a number of circular plates stacked parallel to each other so as to give a large effective surface. The plates are of graphite (a material which does not become highly radioactive under neutron bombardment) covered with a layer of boron enriched with isotope B^{10}. The latter has a large effective cross-section for thermal neutrons. When a B^{10} nucleus absorbs a neutron, an alpha particle ($_2He^4$) is liberated:

$$_5B^{10} + {}_0n^1 \rightarrow {}_2He^4 + {}_3Li^7,$$

and these alpha particles as well as the lithium nuclei cause ionization in the nitrogen with which the chamber is filled. In this way an ionization current is produced which is proportional to the thermal neutron flux.

A simplified circuit diagram of the safety amplifiers is reproduced in *fig. 11*. The anode circuit of a thyratron Th contains a relay coil A and is supplied with a square-wave voltage of frequency 2000 c/s. Through resistor R_1, which connects the grid with the cathode, a current I flows from the ionization

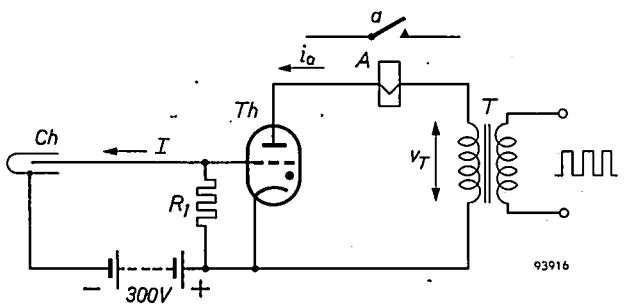


Fig. 11. Simplified circuit of safety amplifier. The current I from the ionization chamber Ch produces across resistor R_1 a voltage which, when greater than a critical value, prevents the thyratron Th from igniting. The relay A is then de-energized and contact a opens, the result of which is to stop the reactor. The anode circuit is supplied with a square-wave voltage via transformer T.

[4]) At the exhibition "Het Atoom" a safety unit of American manufacture was used. In Delft this will be replaced by a PW X086 unit.

chamber *Ch*. The grid voltage is thus a direct voltage proportional to I, hence proportional to the neutron flux at the position of the ionization chamber. As long as I is smaller than a critical value, the thyratron passes a constant current during half of each cycle — owing to the square-wave form of the supply voltage (*fig. 12*) — and this current actuates the relay. If I exceeds the critical value, the grid voltage goes so negative that the thyratron can no longer ignite, and consequently the relay cuts out; the contact a (fig. 11), through which the current for the electromagnets passes, then opens and the safety rods fall into the reactor.

In all there are four identical safety amplifiers, two of which are connected to one ionization chamber and two to the other, as indicated in *fig. 13*. In this figure one of the amplifiers is shown in more detail. The thyratron, type PL 2D21, has a screen grid, which is connected to the cathode. The relay coil is shunted with a capacitor C for by-passing the alternating-current component of the anode current. G is the rectifier which supplies the electromagnets M. The relays A', A'', ... have two sets of contacts, viz. a_2', a_2'', ..., which interrupt the direct current, and a_1', a_1'', ... which, as an extra precaution, cut out the rectifier at the alternating-current side. Each relay also has a contact (not shown) which causes a red lamp to light up on the control desk for as long as the relay is not energized.

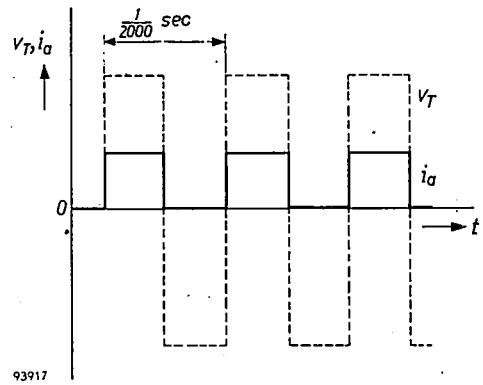


Fig. 12. Supply voltage v_T and anode current i_a of the thyratron in fig. 11, as a function of time t.

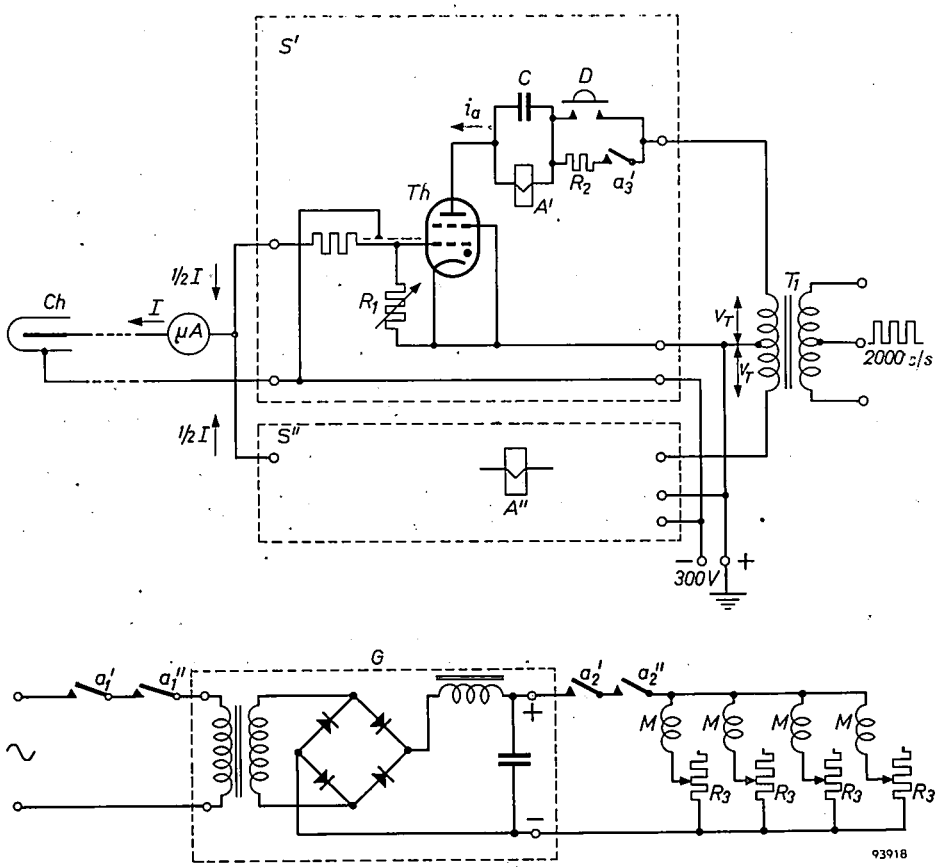


Fig. 13. Two identical safety amplifiers (cf. fig. 11), S' and S'', are connected to one ionization chamber *Ch*. R_1 is a variable resistor for adjusting the current I at which the safety system operates. *Th* thyratron PL 2D21. A' relay type T 51 C, with contacts a_1', a_2' and a_3'; the contacts a_1'' and a_2'' belong to the corresponding relay A'' in S''. C capacitor (A.C. shunt). R_2 series resistor, short-circuited by push-button D. T_1 power transformer which supplies a square-wave voltage of 2000 c/s to both amplifiers. The rectifier G supplies the four electromagnets M via variable resistors R_3.

The current I at which the amplifiers enter into operation (the "scram point") can be adjusted by variable resistors R_1. These resistors also serve to correct for the spread in the characteristics of tube PL 2D21.

After the reactor is stopped, the current I becomes zero. The thyratrons would now be able to ignite again if their anode circuits were not interrupted by the contacts a_3', ... , which are also opened when the relays cut out. In order to actuate the relays again, push-button D is pressed for a moment, so that contacts a_3', ... are shunted. Since this also short-circuits the resistors R_2, anode currents flow which are somewhat higher than normal, so that the relays cannot fail to be actuated. When D is released, the resistors R_2 are switched in again and the circuit returns to normal.

A photograph of the safety unit is given in *fig. 14*.

always a chance of faults occurring which do not result in complete shut-down. Safety amplifiers should be so designed as to reduce the risk of such faults to a minimum. On the other hand it is of the utmost importance to avoid *unnecessary* stopping of the reactor: it is a nuisance and time-consuming in the case of a research reactor but can have serious economic consequences in the case of a power reactor. The safety system must therefore reach a very high standard of *reliability*. This means that only components of the highest quality should be used, and that the apparatus should be as simple as possible.

Safety can also be improved by duplicating the installation; in our case two ionization chambers are employed and four safety amplifiers. Naturally, this slightly increases the chance of unnecessary stoppage.

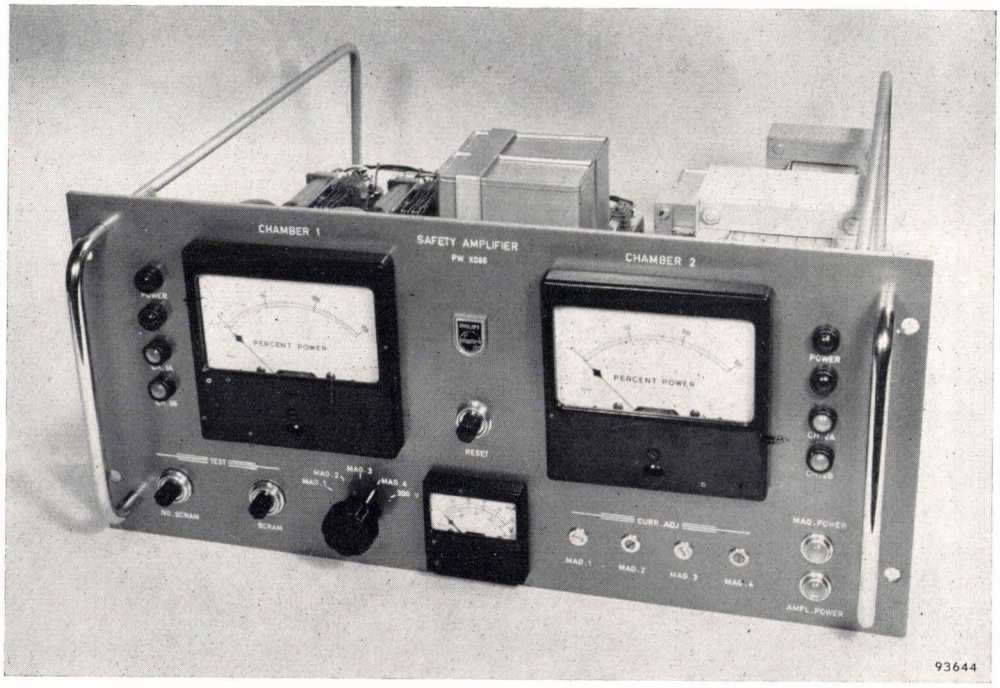


Fig. 14. Unit PW X086 with four safety amplifiers as in fig. 13. The two large meters indicate the current through the ionization chambers and have a relative scale for reading off the percentage power. Between them is a push-button (D in fig. 13) for resetting the amplifier after stoppage. Below: meter for testing the supply voltage and magnet currents, with five-position switch. Further to the left are the test push-buttons for feeding different calibrating currents to the input. Right, screws for adjusting the four resistors to regulate the magnet currents (R_3 in fig. 13).

"Fail-safe" measures and reliability

An important requirement imposed on a safety amplifier is that, no matter what internal defect arises, safety should not be jeopardized, i.e. any fault in the safety equipment itself must result in the reactor being shut down. In reality this "fail-safe" requirement can never be entirely fulfilled; there is

Closer examination [5]) shows that the chances of failure are multiplicative and the chances of unnecessary stoppage additive. Thus, the situation is always relatively improved by duplication.

In the U.K., for example, a system has been adopted with three safety systems, on the recommendation of the Atomic

[5]) E. Siddall, Reliable reactor protection, Nucleonics **15**, No. 6, 124-129, 1957.

Energy Research Establishment at Harwell; these are so arranged that when only one of them comes into operation, merely an alarm is sounded, and the reactor is stopped only when two (or all three) operate. With this system the chance of unnecessary stoppage is much smaller than with a duplicated installation, while in both cases the chance of failure is of the same order.

In *Table I* a summary is given of defects that may occur in the safety amplifiers in unit PW X086 and of their consequences. From the notes appended to this table, it can be seen that the amplifiers are to a high degree "fail-safe". The fact that this result is achieved with relatively few components adds substantially to the reliability of the system.

A further measure adopted to prevent failure of the amplifier is that the screening of the control-grid leads is not connected to earth but to a point of —300 V potential (fig. 13); this does not detract from the screening, since the source of 300 V has a low impedance, and it has the advantage that the relay is de-energized in the event of a short-circuit between a point of the circuit and the screening.

The advantage of using a common direct-voltage source for the ionization chambers and the square-wave generator is that failure of this source also de-energizes the (relay the square-wave voltage having disappeared). As can be seen from fig. 13, no point in the circuit has a positive potential with respect to the cathode; there is therefore no possibility of a positive direct voltage appearing on the anode as a result of a short-circuit — which would keep the thyratron ignited and the relay energized.

Another important point worth mentioning is the deliberate use in the circuit of a rectifying element (the thyratron) enabling the D.C. relay to be energized from a source of alternating voltage. In the event of a short-circuit between anode and cathode, or between anode and screen grid, the relay will therefore be de-energized.

As stated, the safety amplifiers are fed by a square-wave voltage, the reason being that the anode current pulses then always last for half a cycle and thus the average anode current is independent

Table I. Possible defects in the safety amplifiers, and their consequences.

Resistor R_1 interrupted	relay de-energized
„ „ short-circuited	relay stays energized
μAmmeter interrupted *)	„ „ „
„ „ short-circuited	no effect
Relay coil open-circuit	relay de-energized
„ „ short-circuited	„ „
Capacitor C interrupted	„ „
„ „ short-circuited	„ „
Resistor R_2 interrupted	„ „
„ „ short-circuited	„ „ **)
Thyratron: leak	„ „
„ : interruption filament current	„ „
„ : „ control grid (g_1)	„ stays energized
„ : „ screen grid (g_2)	„ de-energized
„ : „ anode (a)	„ „
„ : short-circuit cathode (k)-g_1 ***)	„ stays energized
„ : „ „ k-g_2	no effect
„ : „ „ k-a ***)	relay de-energized
„ : „ „ g_1-g_2	„ stays energized
„ : „ „ g_1-a ***)	„ „ „
„ : „ „ g_2-a	„ de-energized
Failure of 2000 c/s generator	„ „
„ „ 300 V supply	„ „

*) The chance of interruption is very slight, since the meter is fitted with a shunt.

**) Since, with short-circuited R_2, the anode current is higher than usual, the relay opens more slowly and only at a higher ionization current I.

***) The construction of tube PL 2D21, and the wiring are such that short-circuiting of k-g_1, k-a and g_1-a is practically impossible.

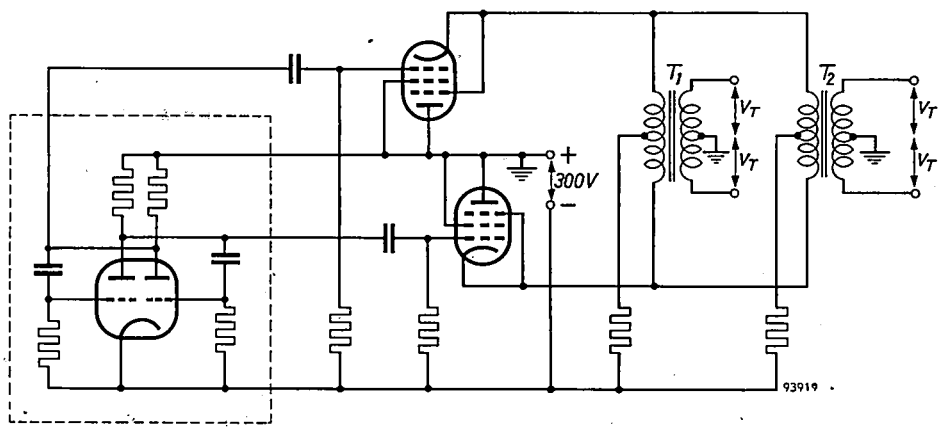


Fig. 15. Circuit diagram of 2000 c/s square-wave generator. Inside the broken lines is the multivibrator with double triode E 90 CC. Right, balanced push-pull output stage with two pentodes EL 34 giving cathode-follower output. T_1 and T_2 are the output transformers, each of which supplies two safety amplifiers.

of the grid voltage. The frequency is chosen as 2000 c/s because one complete cycle can elapse before the relay current is interrupted, 2000 c/s being equivalent to 0.5 millisecond. Added to the 4 milliseconds opening time of the relay (type T 51C), this gives a total time of 4.5 milliseconds. "Less than 5 milliseconds" is the requirement; this time is so short that the falling rods, even in the most unfavourable case, adequately slow down the reaction.

Finally, some words on the test device and on the square-wave voltage generator.

Test device

The operation of the safety system can be tested as follows. Pressing a button causes a specific direct current to be fed to the input of the safety amplifier; when this is done the relay should become de-energized. By pressing another button a current about 10% lower is fed to the input; the relay should then stay energized.

Both currents are taken from the direct-voltage source of 300 V and can be adjusted with a variable resistor inside the unit.

The square-wave voltage generator

The square-wave voltage of 2000 c/s is supplied by a generator (*fig. 15*) consisting of a multivibrator and a power amplifier, both fed by a direct voltage of 300 V. The multivibrator (with a double triode E 90 CC) delivers a square-wave voltage (substantially independent of the valve characteristics) of 140 V (280 V peak-to-peak, hence almost the maximum obtainable with 300 V supply voltage). The latter voltage drives a balanced push-pull output stage containing two EL 34 pentodes with cathode follower output. The efficiency of this stage is low, but it has the advantage that variations in tube characteristics have hardly any effect on the voltage gain (which is approximately 1).

There are two output transformers (T_1, T_2), whose primaries are connected in parallel. Each has two secondary windings and supplies two safety amplifiers (see fig. 13). A tight coupling between primary and secondary coils precludes overshoot transients.

Summary. The article describes three component units of the monitoring equipment for the nuclear reactor displayed at the exhibition "Het Atoom" and shortly to be installed at the Technische Hogeschool, Delft. These units are:

1) The logarithmic measuring channel for high neutron flux. Besides an ionization chamber this comprises: a logarithmic amplifier, in which the logarithmic element is a triode operating in the exponential region; a "period amplifier", which supplies a voltage inversely proportional to the period of the reactor; a relay stage, with two relays which open safety contacts when the period drops to 10 and 3 seconds respectively; and various calibrating devices.
2) The gamma indicator, with four ionization chambers of 800 cm^3 volume, filled with argon at 15 atm. The ionization current is measured with an electrometer tetrode 4066, which is mounted inside the ionization chamber to avoid leakage currents. A light-relay utilizing a transistor is actuated by this tube and so trips an ordinary relay.
3) The safety amplifiers, which are designed with "fail-safe" features, i.e. to ensure that the reactor is stopped in the event of any defects in the safety equipment itself; in this way defects in the safety equipment have practically no effect on the safety of the whole, while unnecessary stoppages are avoided as far as possible.

DIRECT OBSERVATION OF WEISS DOMAINS BY MEANS OF THE FARADAY EFFECT

538.614

By cooling a nearly eutectic melt, large plate-like single crystals of the hexagonal compound $BaFe_{12}O_{19}$, the main constituent of ferroxdure, can be obtained. In this process, now in use for some years in the Philips Research Laboratories [1]), extremely thin crystal plates may be obtained under certain conditions. Their thickness is less then a micron and their surface area of the order of one square millimetre. X-ray diffraction photographs of these tiny platelets have shown that they too are single crystals of $BaFe_{12}O_{19}$, with the hexagonal axis perpendicular to the flat surface.

[1]) A. L. Stuijts, to be published elsewhere.

The special interest of these platelets is that they are reasonably transparent to red light. (This is related to the high resistivity of $BaFe_{12}O_{19}$, a characteristic which is due to the fact that all iron ions present in this compound have valency 3. The presence of even small quantities of iron ions of valency 2 besides those of valency 3 will in general cause a much lower resistivity and a strong absorption of light.) The transparency of the platelets opens up an elegant method for the direct observation of the Weiss domains in $BaFe_{12}O_{19}$. The platelet is observed in a polarizing microscope with nearly-crossed nicols. One then sees an image such as those depicted here in *figs. 1-3*.

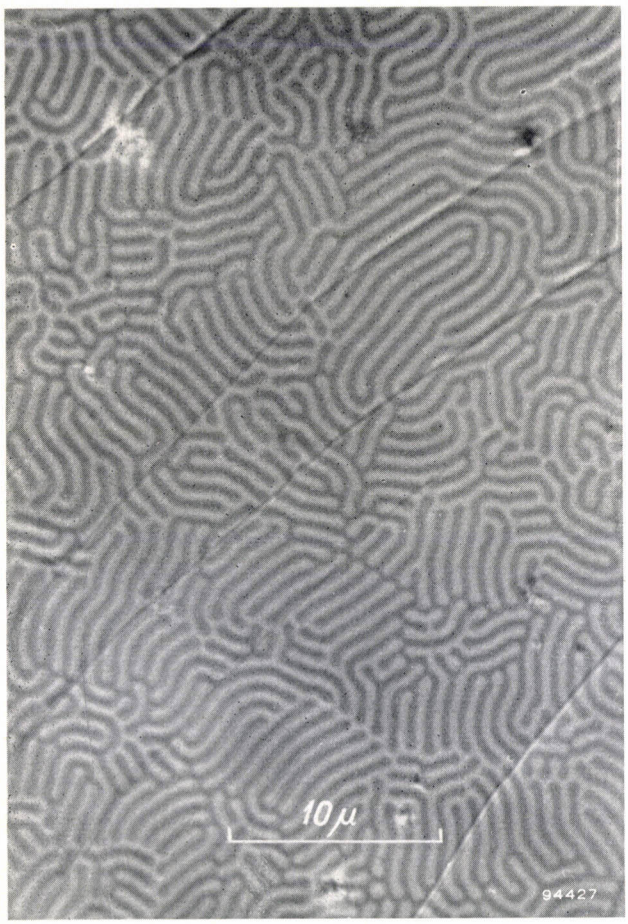


Fig. 1.

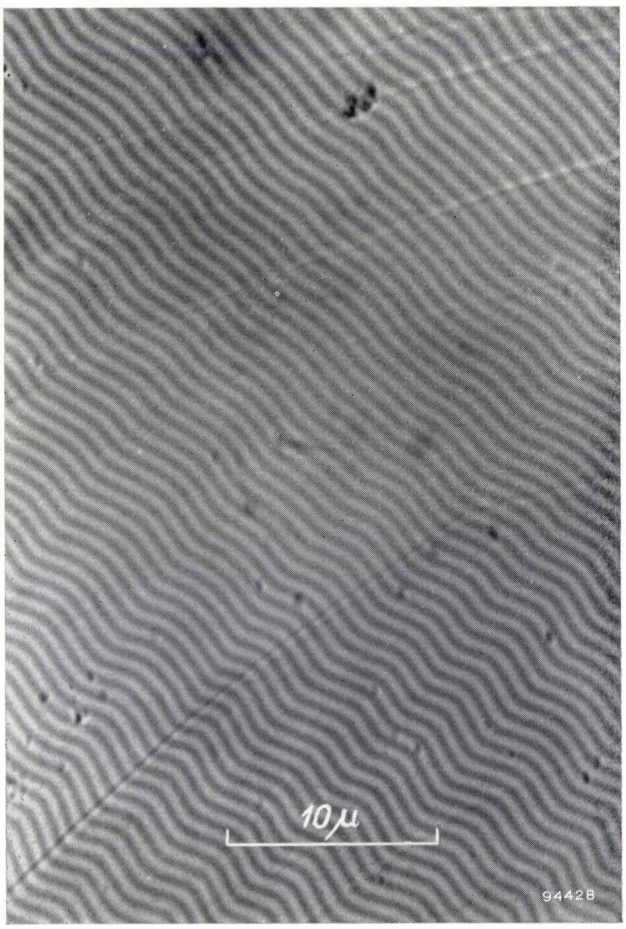


Fig. 2.

Fig. 1. Single-crystal platelet ($< 1\ \mu$ thick) of $BaFe_{12}O_{19}$ in the virgin state, as seen under a polarizing microscope with the nicols not quite exactly crossed. The light and dark bands are elongated Weiss domains.

Fig. 2. The same platelet as in fig. 1, after magnetizing. The domain pattern has assumed a much simpler form. (From scratches and spotty surface faults one can see that this photo represents almost the same region as that in fig. 1.)

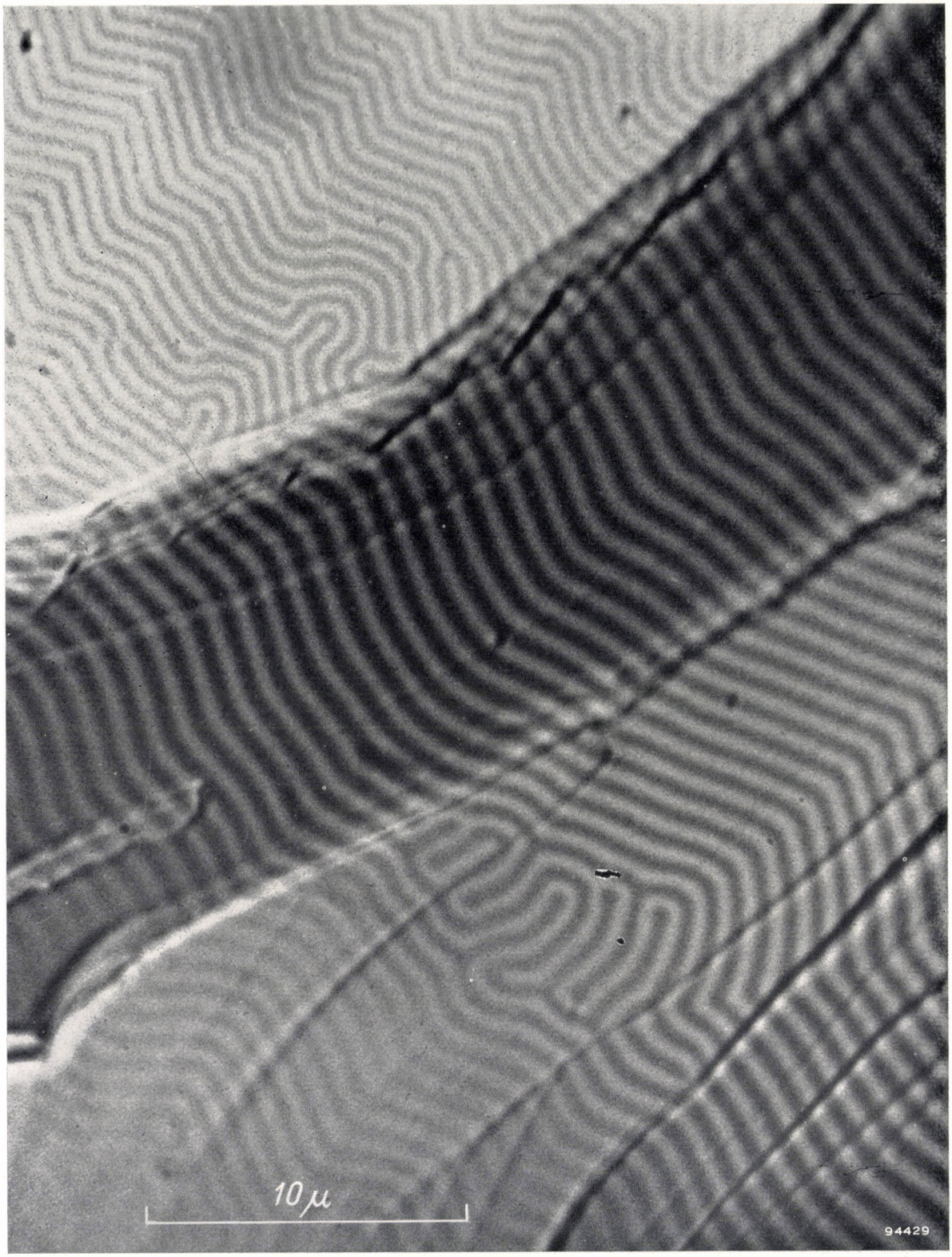


Fig. 3. Similar photo to that of fig. 2, showing another region of the same platelet. Terraced thickening of the specimen can be seen in this region by the dark broad bands across the pattern. The Weiss domains are seen to be wider within these dark bands. The "contraction" of the pattern on passing the boundary from a thick to a thinner place is at some points taken up by dislocation-like imperfections in the pattern.

A brief explanation of the occurrence of these patterns and an indication of some of the details to be seen may be appropriate here.

In each Weiss domain in a ferromagnetic material the magnetic moments of the atoms or ions are parallel aligned, so that in each domain the satura-

tion magnetization exists. The direction of magnetization, which depends on the crystal-shape- or stress-anisotropy, differs from one Weiss domain to another. In $BaFe_{12}O_{19}$ the direction of the magnetization is entirely determined by the very large crystal anisotropy; the preferred direction of magnetization is here the hexagonal axis (*c*-axis). In the crystal platelets therefore, the magnetization of all Weiss domains is perpendicular to the flat surface, in spite of the high demagnetization factor of the plate form. The Weiss domains are thus magnetized as sketched in *fig. 4*, with the domain walls all perpendicular to the flat surface. Because the magnetization is so strongly bound to the hexagonal axis in $BaFe_{12}O_{19}$, no "closing" domains occur at the surface (i.e. Weiss domains with magnetization parellel to the surface), in contrast with the usual situation in other materials.

If a beam of plane-polarized light is now projected parallel to the hexagonal axis on to such a platelet, a Faraday rotation occurs, of magnitude and sign dependent on the magnetization. Suppose that for one domain the plane of polarization undergoes a rotation described by a right-handed screw; the light passing through an adjacent domain, owing to the oppositely-directed magnetization, will then undergo an equal rotation of its plane of polarization in the opposite direction, i.e. corresponding to a left-handed screw. An analysing nicol is placed in the emerging beam. By placing this not exactly perpendicular to the plane of polarization of the incident beam, the intensity of the rays passing through adjacent, oppositely-magnetized domains will differ slightly; see *fig. 5*. In this way the whole domain pattern becomes visible [2]).

The pictures shown in figs. 1-3 were made in this manner. The hexagonal axis of the crystal lies perpendicular to the plane of the photographs. The Weiss domains, with their magnetizations directed alternately upwards and downwards out of the plane of the picture, are seen as light and dark bands. The remarkable labyrinth structure of fig. 1 is the domain pattern in the *virgin state*, as occurs after cooling the material from a temperature above the Curie point. Fig. 2 is a photo of practically the same region as that in fig. 1, taken after magnetizing the specimen in a field of 48×10^4 A/m [6000 oersteds] parallel to the hexagonal axis. The domain pattern has here assumed a much simpler form; the ribbon-like domains are seen as wavy lines. The remanent magnetization of the platelet is practically zero as a result of the very large demagnetization factor

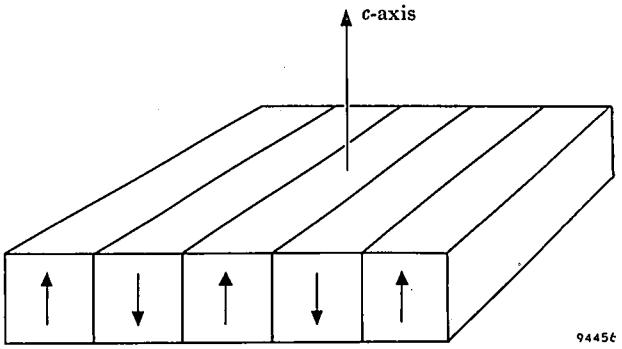


Fig. 4. Schematic diagram showing the configuration of Weiss domains in the single-crystal platelets of $BaFe_{12}O_{19}$. The domain walls are perpendicular to the flat surface of the platelets, parallel to the hexagonal *c*-axis. The arrows show the directions of the magnetization.

with such a thin specimen. This can be seen in the photo since the light and dark bands are equally wide; in the total volume, therefore, their is no surplus of one direction of magnetization over the other.

The *width* of the Weiss domains depends on the thickness of the crystal plate. The formation of the domain walls requires a certain energy, proportional to the wall area; on the other hand the magnetostatic energy of the whole body is reduced by the formation of the domain walls. With increasing thickness of the plate the magnetostatic energy becomes smaller, whereas the domain walls would take up more energy owing to their greater area (greater height) if their spacing remained the same. The total energy

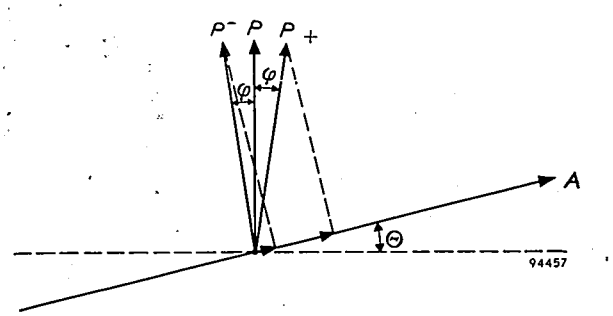


Fig. 5. If P is the direction of polarization of a light beam (here considered incident perpendicularly to the paper), the Faraday effect in one Weiss domain gives rise to a rotation φ to $P+$, while in adjacent oppositely-magnetized domains the same rotation occurs in the opposite sense, to P^-. An analyser set with its transmission direction at A, $90° - \Theta$ from P, attenuates the light with the polarization P^- more than that with the polarization $P+$.

The transmitted intensities are related to each other as $\sin^2(\Theta - \varphi) : \sin^2(\Theta + \varphi)$. Maximum contrast is obtained at $\Theta = \varphi$, but for a small value of φ the transmitted intensity is too small when $\Theta = \varphi$. One therefore takes a somewhat larger value of Θ at which the image gives a good compromise between intensity and contrast.

[2]) For methods for the *surface* observation of Weiss domains (Bitter patterns, magneto-optical Kerr effect, etc.) see, for example, R. M. Bozorth, J. Phys. et Rad. **12**, 308, 1951 and B. W. Roberts, Conf. Magnetism and Mag. Mat., Pittsburgh, Pa., 14-16 June 1955, p. 192.

The method using the Faraday effect has also been used by P. A. Miles, who employed iron oxides of garnet structure as the transparent ferromagnetic (Prog. Rep. M.I.T. Lab. Insul. Res. **21**, 17, 1957).

is therefore minimized by a reduction of the number of domain walls per cm length, i.e. the domains become broader. This effect is strikingly demonstrated in fig. 3. This photo is analogous to fig. 2 but represents another region of the specimen. The platelet has here a terraced change in thickness which can be seen (owing to the increased absorption) as broad dark bands passing through the whole domain pattern. Inside the dark bands the Weiss domains are seen to be broader than in the lighter regions.

The elongated Weiss domains run continuously through the light and dark regions and apparently are not affected by the "steps" on the surface. To enable this continuity to occur, with a different number of domains per cm length in adjacent regions, some geometrical rearrangement is necessary. One can see how this occurs in the pattern; at the boundary of *A* and *B*, "extra" Weiss domains can be seen to end at various places on the boundary, showing a certain resemblance to edge dislocations in a crystal lattice. At one place on the boundary of *B* and *C* a region of meandering Weiss domains occurs, which serves to fill in a space which has arisen owing to the contraction of the domains in *C*.

Finally, a remark concerning the Faraday effect which is responsible for making the domain pattern visible. The Faraday effect in ferromagnetics has been known for some years, viz. in ferrites which are transparent to electromagnetic waves of centimetre and decimetre wavelengths. From the theory of the propagation of these waves in ferromagnetics, due to Polder [3]), a rotation of the plane of polarization is to be expected in the presence of a suitable external magnetic field; this is a result of the interaction between the precessing electron spins (aligned by mutual interaction) and the rotating *magnetic* field of a circularly polarized wave with the right sense of rotation. This is the same interaction as causes ferromagnetic resonance. The Faraday effect in ferrites has been usefully applied in ferrite isolators for microwaves [4]). One is tempted to attribute the Faraday rotation in $BaFe_{12}O_{19}$ at optical wavelengths, with which we are concerned here, to the same mechanism as that in ferrite isolators at microwave frequencies. This is not borne out quantitatively, however. A calculation of the rotation to be expected on the basis of magnitude of the spin-field interaction (known from other effects), results in a value about 100 times smaller than that observed. The Faraday effect in the present case must in fact be attributed to the interaction between the orbital electrons (whose orbits are aligned by the spin-orbit interaction in a ferromagnetic) and the rotating *electric* field of the circularly polarized light beam [5]). This is essentially the same mechanism as that responsible for the normal Faraday effect with light waves in diamagnetics. In the ferromagnetic $BaFe_{12}O_{19}$ the effect is so strong that even a thickness of 1 micron provides an observable effect.

C. KOOY.

[3]) D. Polder, On the theory of ferromagnetic resonance, Phil. Mag. **40**, 99-115, 1949. See also H. G. Beljers and J. L. Snoek, Gyromagnetic phenomena occurring with ferrites, Philips tech. Rev. **11**, 313-322, 1949/50.

[4]) C. L. Hogan, The ferromagnetic Faraday effect at microwave frequencies and its applications. The microwave gyrator; Bell Syst. tech. J. **31**, 1-31, 1952.
A. G. Fox, S. E. Miller and M. T. Weiss, Behavior and applications of ferrites in the microwave region, Bell Syst. tech. J. **34**, 5-103, 1955. See also H. G. Beljers, The application of ferroxcube in unidirectional waveguides and its bearing on the principle of reciprocity, Philips tech. Rev. **18**, 158-166, 1956/57 (No. 6).

[5]) H. R. Hulme, Proc. Roy. Soc. A **135**, 237, 1932.

BARIUM GETTER FILMS

by J. J. B. FRANSEN and H. J. R. PERDIJK.

621.385.1.032.928

During the many years of mass-production of electron tubes, a great deal of experience has been accumulated regarding the best methods of depositing the barium getter films, which serve to bind any residual or liberated gases. In spite of much research, however, it has hitherto not been entirely clear why one method is more efficient than another.

In the article below the authors put forward a hypothesis, based on a variety of experiments, concerning the structure of the barium getter films in most commercial valves and the manner in which the films occur. The hypothesis is very simple and makes it possible to give a straightforward explanation of the experimental results.

In electronic valves there must be no increase in the gas pressure during operation. The gases liberated in a radio valve cause the grid current to rise and can also lead to undesirable chemical reactions. A chemical reaction with the cathode can result in a drop in emission, and a reaction with the substances used for coating the grids changes the characteristics of the valve. Adsorption, too, at the surface of a grid can so alter the condition of the grid as to affect the valve characteristics. In a gas-filled valve any contamination of the gas will alter the ignition, burning and extinction voltages.

In order to limit as much as possible the increase of gas pressure in a sealed-off electronic valve, the components are subjected to a suitable pre-treatment, and during evacuation the valve is outgassed by heating to a high temperature, perhaps combined with electron bombardment. This does not, however, prevent the liberation of gases entirely. The gases in question are found to be mainly carbon monoxide and hydrogen, and to a lesser extent water vapour. In certain cases nitrogen, oxygen and carbon dioxide may occur [1]).

For the removal of these gases all valves are provided with an "internal pump", which goes on functioning during the whole useful life of the valve. This pump consists simply of a certain quantity of a substance (the getter material) which is capable of absorbing gases (usually a chemical binding is involved). To ensure that the process takes place quickly and in such a way that as much as possible of the getter material really remains active (it might, for example, become inactive merely owing to inaccessibility by being blocked off), it is usually necessary for the material to be applied in a special manner inside the valve envelope.

The substance first used for "gettering" was phosphorus. It is still used for this purpose in incandescent lamps, but in radio valves its place was soon taken by magnesium. Nowadays the getter materials most commonly used are barium (for radio valves and television picture tubes) and zirconium (for transmitting valves).

In this article we shall confine ourselves to barium. Before going into details concerning the technique of using barium as a getter material, it should be mentioned here that the metal is present in the finished valve or tube in the form of a thin film deposited on a part of the inside wall. With reference to some experiments which we have carried out we shall try to give a picture of the physico-chemical background of the rules, familiar in practice, governing the application of the barium film. The hypothesis underlying this picture is confirmed by electron-microscopic examination. It will also be shown what chemical reaction is, in all probability, responsible for the absorption of carbon monoxide.

First of all we shall consider some aspects of the techniques relating to barium getters.

The use of barium as a getter material

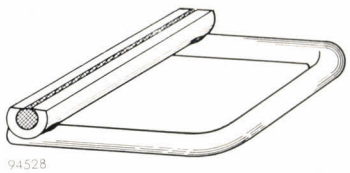
In the early period of barium getters the free element was usually introduced by suction into thin tubes of nickel, short lengths of which were mounted at an appropriate position inside the valve and heated after evacuation of the bulb. The barium vaporized and formed a thin deposit (the getter film) on the glass wall. This technique has now practically fallen out of use, mainly because the "packing" — the so-called getter holder — does not adequately protect the barium from spoiling in the time between its production and its employment in the valve.

[1]) Cf. S. Wagener, Vacuum **3**, 11, 1953, where the various gases given off by different parts of a valve are mentioned. See also J. Peper, Philips tech. Rev. **19**, 218-220, 1957/58 (No. 7-8).

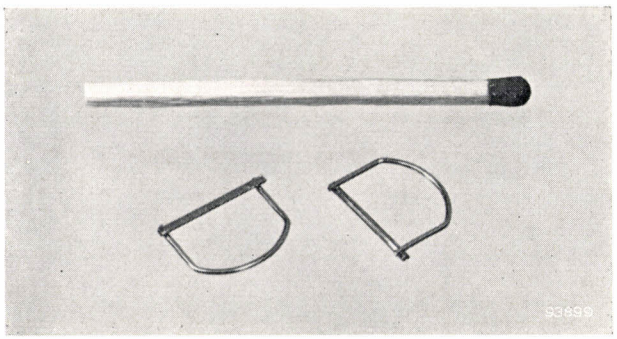
Getter holders were subsequently developed containing a certain quantity of a barium compound which is not affected by contact with the air. At the present time the most widely used material is a barium-aluminium alloy in the proportion of one barium atom to four aluminium atoms. We shall therefore be concerned solely in this article with getter films originating from this alloy.

Types of getter holders

The form of a getter holder differs according to the amount of barium to be vaporized. If only small amounts are required, the holder is usually an iron tube with a top segment cut away and with a "stirrup" of nickel wire welded to both ends (see *fig. 1*). The closed circuit thereby produced enables the getter holder to be heated by means of high-frequency induction. The heating gives rise to a reaction in which the aluminium forms a bond with the iron and the barium escapes.



a



b

Fig. 1. *a*) Illustration of a getter holder having the form of a longitudinally ground-off tube. The "stirrup" welded to it makes it possible to heat the holder by induction.
b) Photograph of two such getter holders.

If greater quantities of barium are required, metal strips are used in which round depressions have been made at equal distances apart. Each depression is filled with a "pill" of the material supplying the barium. Short pieces of such strips, containing, say, two pills, can be made suitable for inductive heating by fitting them with loops as in the case of the iron tubes; longer strips can be bent into the form of a ring (*fig. 2*). In *figs. 3a* and *4* can be seen how the getter holders commonly used in a radio valve and television picture tube, respectively, are mounted on the electrode system.

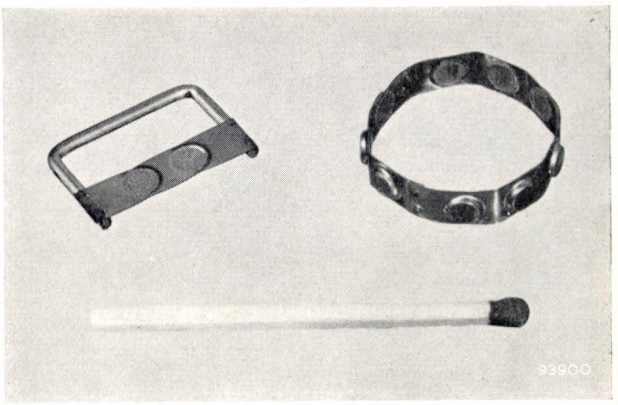
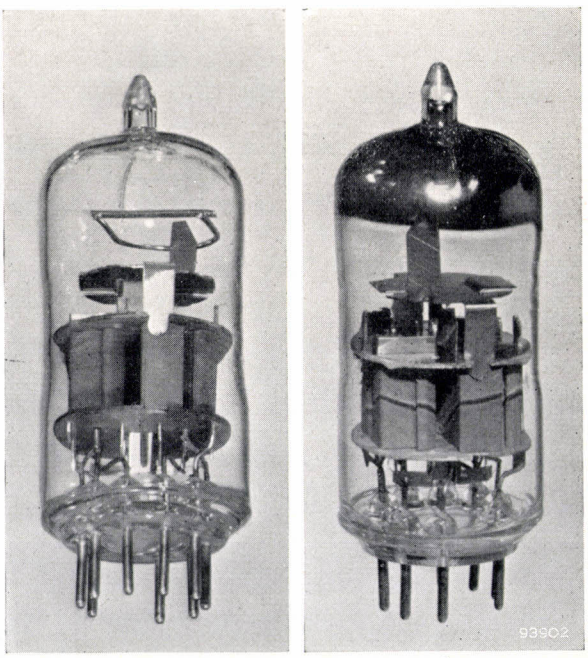


Fig. 2. Getter holders made from a metal strip with round equally spaced depressions containing barium "pills". For the purpose of induction heating a loop is welded to the getter shown on the left. The other consists of a longer strip bent to form a ring.

Vaporizing of the barium; "melt" curve

The barium that escapes from the heated getter holder settles on the neighbouring parts of the bulb wall. The resultant barium film is usually glossy in appearance and is about 1 μ thick. The film can be seen in fig. 3*b* which shows the same valve as in fig. 3*a* after vaporization of the barium.

The following points arise in connection with heating the getter holder. If a high power is applied, there is a risk that the holder will melt before sufficient barium has vaporized. On the other hand, if the h.f. power is only slightly higher than that necessary to start vaporization, quite small varia-



a *b*

Fig. 3. Radio valve containing a getter holder of the type shown in fig. 1, *a*) before, and *b*) after vaporization of the barium.

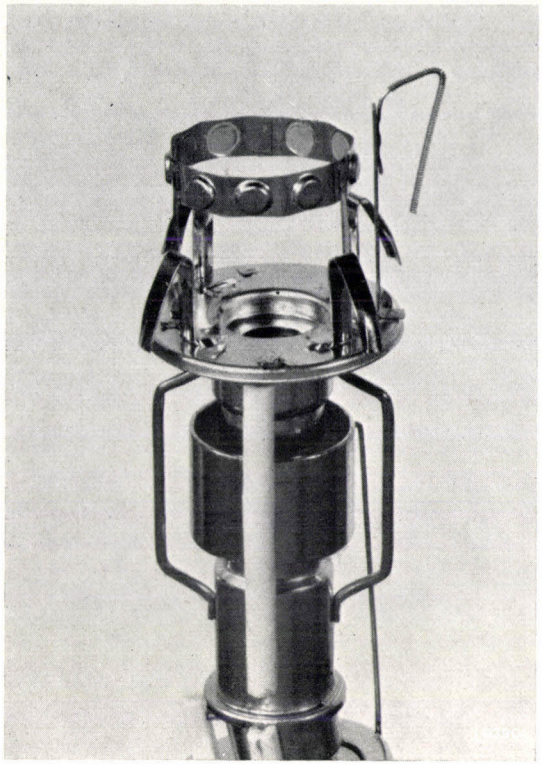


Fig. 4. Part of the electrode system of a television picture tube on which a ring-type holder (fig. 2) is mounted. To avoid any disturbing effect on the electron beam, the getter holder in television picture tubes is made of non-magnetic material.

tions both in the adjustment of the high-frequency generator and in the properties of the getter holder plus filling, will cause large variations in the quantity of barium vaporizing in a given time. Moreover, the barium will then, of course, vaporize very slowly, so that the process must be protracted. During this time the getter holder radiates considerable heat, which raises the temperature of the glass bulb; it will be shown below why this is not desirable.

The manner of heating can be represented by a point in a graph, the ordinate being the time t at which the first deposited barium becomes visible (this is a measure of the power supplied) and the abscissa the time T during which the getter holder is heated. In order to ascertain quickly whether a given combination of h.f. power and heating time is permissible having regard to the risk of melting, it has been found useful to plot a characteristic curve for each type of getter holder. This "melt" curve, as it may be called, is obtained as follows. Samples of the getter holder in question (barium-filled) are each heated by a different h.f. power, and a measurement is made of the above-mentioned time t and the value τ of the time of heating after which the holder melts. The behaviour of each sample is then marked by a single point in the graph, with τ on the abscissa and t on the ordinate (*fig. 5*). The envelope of the points having the higher values of t and the lower values of τ is then drawn, the resultant curve being the melt curve.

It is evident that working-points (combinations of t and T), so chosen that no melting occurs, will as a rule appear in the graph to the left of the melt curve. With the aid of such a graph one can quickly decide what the conditions should be to get the largest possible (and the least variable) fraction of the available barium to evaporate in a short time.

Absorption capacity and gettering rate

For defining the performance of a getter the terms absorption capacity and gettering rate are used.

The absorption capacity is the maximum quantity of gas that a given getter (here the barium film) is capable of absorbing.

The gettering rate a is defined by the formula:

$$\frac{dQ}{dt} = p\,a, \quad \ldots\ldots\ldots \quad (1)$$

where the differential coefficient on the left side represents the quantity of gas absorbed (gettered) per unit time, and p is the gas pressure. The value of a depends on the nature of the absorbed gas as well as on the quality of the getter. It is customary to express p in microns Hg, Q in litre.microns [2]) and t in seconds. The gettering rate a is therefore expressed in litres per second and the absorption capacity, of course, in litre.microns (l.μ).

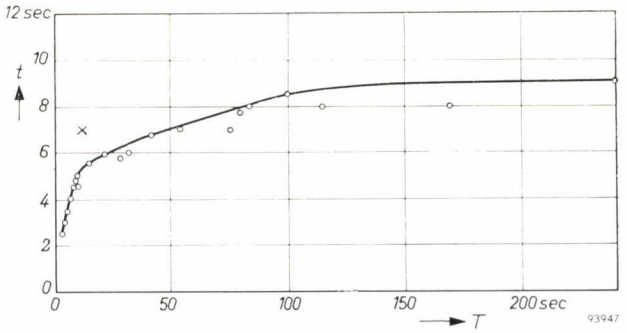


Fig. 5. "Melt" curve for the Philips getter holder R1 679 51. The time t at which the first barium deposit begins to form is plotted against the value τ of the heating time T at which the holder melts through. Each measurement is represented by one point, with τ on the abscissa and t on the ordinate. The higher the power applied, the smaller are t and τ. The envelope of the points representing the lower values of τ, is the "melt" curve. The graph gives the maximum permissible duration of heating at a given t (i.e. at a given applied power). The working-point for the experiments described below with this type of getter holder is marked by a cross in the figure ($t = 7$, $T = 11$ sec).

[2]) The comparison of quantities of gas solely be means of the product of volume and pressure is, of course, only feasible if the temperature remains the same. The numerical results reported in this article refer to room temperature.

It has been found that, to ascertain the quality of a barium getter film, it is necessary to define not only the above quantities, which relate to the whole film, but also corresponding quantities relating to one milligramme of barium and one cm² of the surface covered by the film. These specific quantities are respectively the absorption capacity per mg (unit: l.μ/mg), the absorption capacity per cm² (l.μ/cm²), the gettering rate per mg (l/sec.mg) and the gettering rate per cm² (l/sec.cm²).

Measuring the gettering rate

The set-up which we used for measuring the gettering rate is shown schematically in *fig. 6*. The barium getter film under investigation is contained in the bulb *C*, which is connected via a capillary *W* with the space *A*. The latter is connected via a "lock" *L* (consisting of stopcocks g_1 and g_2) to the gas reservoir *B*, and via a cooler *R* and a stopcock g_3 to the mercury-diffusion pump *P*. The pressure in *A* is measured by a Pirani gauge *M*; the lock *L* makes it possible to feed in a known quantity of gas.

The measurement was carried out as follows. Before forming the film, the spaces *C*, *A*, *R* and *L* were evacuated (g_3 open, g_1 closed). After the film had been formed, g_3 was closed and, by means of lock *L*, a quantity of gas was admitted into space *A*. From this moment onward, the pressure in *A* was measured, the gas present in *A* being able to flow via *W* to *C*, at a speed determined by the resistance of *W* and by the pressure in *C*, that is by the rate at which the getter film absorbed the gas entering *C*.

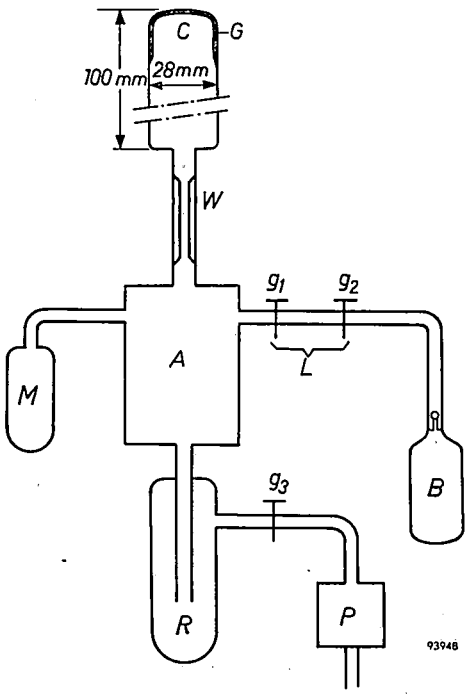


Fig. 6. Set-up used for measuring the "gettering rate" of a barium getter film. A constant quantity of the gas in *B* is admitted to *A* via the lock *L*, consisting of stopcocks g_1 and g_2. Via the capillary *W* the gas flows into the bulb *C* containing the barium getter film *G*. The time-variation of the pressure in *A*, measured by a Pirani gauge *M*, is governed by the permeability of the capillary *W* and the gettering rate to be measured.

Writing the pressures in *A* and *C* as p_1 and p_2 respectively, the total volume of *A*, *R*, *L* and *M* as V_1, the volume of *C* as V_2, the conductivity of *W* as *F*, the time as *t* and the gettering rate as *a*, we can describe what happens in terms of the following equations:

$$V_1 \frac{dp_1}{dt} = -F(p_1 - p_2), \quad \ldots\ldots \quad (2)$$

$$V_2 \frac{dp_2}{dt} = F(p_1 - p_2) - a p_2. \quad \ldots\ldots \quad (3)$$

When $V_1 \gg V_2$, so that V_1 and $(V_1 + V_2)$ can both be put equal to V, and when *a* and *F* are constant, the solution is:

$$p_1 = p_0 \exp \frac{-t}{(1/a + 1/F)V}, \quad \ldots\ldots \quad (4)$$

$$p_2 = \frac{p_0}{1 + a/F} \exp \frac{-t}{(1/a + 1/F)V}, \quad \ldots \quad (5)$$

where p_0 is the initial value of p_1. As stated, the pressure in *A* was measured as a function of time. From equation (4) it follows that the logarithm of the measured p_1 plotted against *t* will produce a straight line. If the amount of gas admitted is sufficiently small, this is indeed found in the course of a single continuous measurement. A family of such curves, obtained from measurements with hydrogen, is given in *fig. 7*. It can be seen that *a* becomes smaller according as more gas has been absorbed by the getter film.

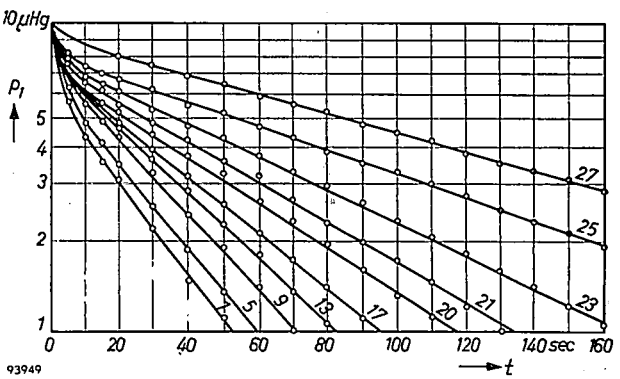


Fig. 7. Gettering rate for hydrogen measured on a single barium film, the logarithm of the pressure p_1 being plotted against the time *t* which has elapsed since the admission of a given quantity of gas. After a short time the relationship is linear. As more gas is absorbed the gettering rate decreases and the curve becomes less steep. The curves represent the pressure gradient after the first, fifth, ninth, etc. admissions of gas.

The quality of a barium getter film in relation to its structure and manner of preparation

The quality of a barium film as a getter depends on many factors and is by no means exclusively determined by the quantity of barium which it contains or by the surface area of the glass wall covered by barium. Besides such factors as the form of the getter holder, the quantity of barium in it, whether or not it has been subjected to prior outgassing and the method of heating it when vaporizing the barium, an important part is played by the temperature of the bulb, the distance of the getter holder from the bulb wall and the presence of gas in the valve during the vaporizing process.

In experimental investigations on the influence of each of the above factors it is, of course, necessary to keep the others as constant as possible. With factors such as the temperature, the means of arranging this are fairly obvious, and no special description is required. Others, such as the shape of the holder, can certainly be kept constant within a series of experiments but the precise conditions are difficult to describe, for which reason they are treated rather cursorily in some publications. The outcome is that the results reported in these publications can hardly be compared quantitatively with others [3]).

In view of the experiments described in this article and in the light of the experience which we and others have gained with barium getter films, we believe that the production of a barium film and its structure should be regarded as taking place in the following way.

When the getter holder, with its barium-containing material, is heated to the temperature for vaporizing the barium, not only does the barium escape but gases are liberated as well, it not being possible to outgas the getter holder completely. There is reason to assume that, in consequence of this, the barium does not fly in atomic form from the getter holder to the bulb wall but rather as a "mist" composed of minute droplets. The diameter of these droplets is determined in the immediate vicinity of the holder.

The mechanism of this process may well be such that fairly numerous collisions take place near the getter holder between gas molecules and barium atoms, whereby droplets of barium condense from the vapour to form the mist described [4]). There are other arguments, however, for the view that the barium is already in the form of droplets upon leaving the holder.

At all events, a barium getter film produced from such a mist will consist of a porous structure of stacked round grains of barium. The "microscopic" surface of this film is larger (per mg barium) according as the barium grains are smaller and less fused together. They will fuse together less, the lower is the temperature of the bulb and of the droplets at the moment of arrival.

Photomicrographs, made with an electron microscope, of the surface of barium getter films support this hypothesis (*fig. 8*). The surfaces of two films containing respectively 0.9 and 2.6 mg barium (over an area of 35 cm^2) are seen to have a grainy structure. The surface of the third film, containing 3.5 mg barium, gives the impression of being formed by grains which have partly fused together. The grains visible on the micrographs of the first two films have a diameter of about 0.04 μ [5]).

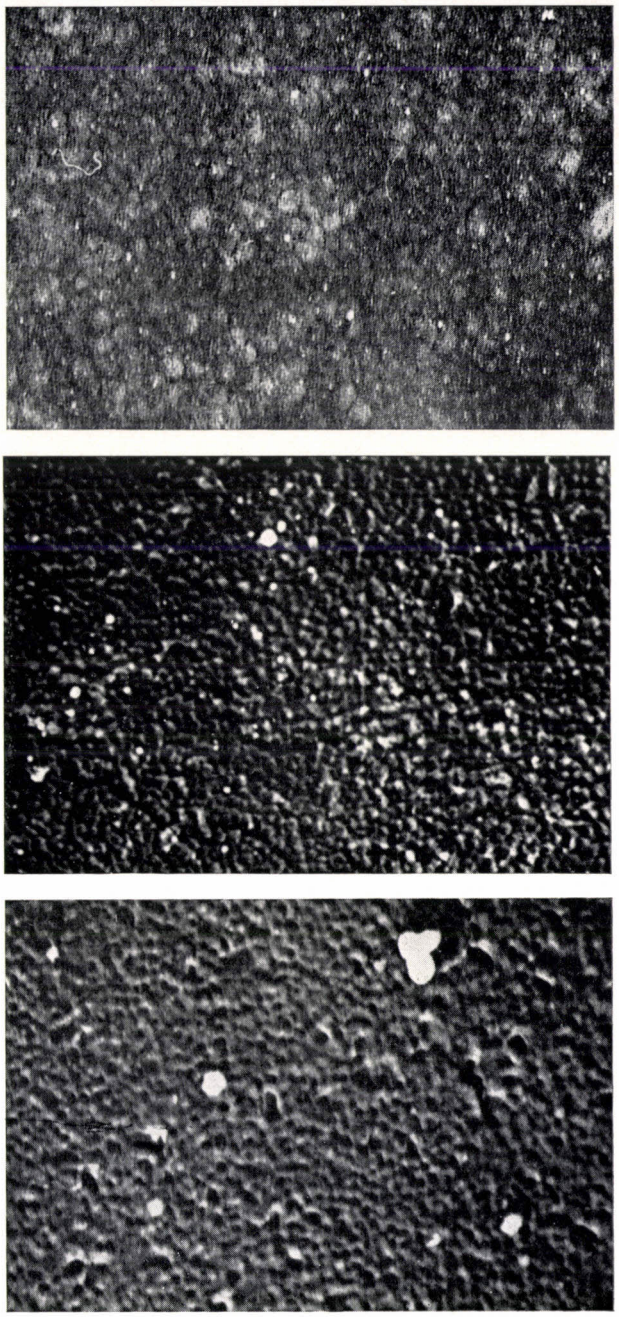


Fig. 8. Photomicrographs made with an electron microscope, of SiO replicas of the surface of barium films having a macroscopic area of 35 cm^2 and containing respectively 0.9, 2.6 and 3.5 mg barium. Magnification 80 000 ×.
a) Surface of film containing 0.9 mg Ba, exhibiting more or less spherical grains with a diameter on the photograph of approx. 3 mm; the actual grain diameter is thus about 0.04 μ.
b) and *c*) Surfaces of films containing respectively 2.6 and 3.5 mg Ba. The grains are still clearly distinguishable, but are now somewhat fused together, particularly in *c*.

[3]) For example, observations by Wagener (Vacuum **3**, 11, 1953) of the absorption of hydrogen conflict with those by Della Porta (Vacuum **4**, 284, 1954). Our results are complementary to those of the former author.

[4]) See J. E. de Graaf and H. C. Hamaker, Physica **9**, 300 and 301, 1942 and Y. Mizushima, Z. Oda and O. Ochi, J. Phys. Soc. Japan **12**, 365, 1957 (No. 4).

[5]) These photomicrographs were made by H. B. Haanstra of Philips Research Laboratories.

For examination with the electron microscope it is, of course, necessary to use a replica technique, such that the replica of the barium film can be formed in the evacuated valve. A method was adopted whereby a thin SiO film was vapour-deposited on the object [6]). The $Si\text{-}SiO_2$ mixture, which was heated to produce the SiO, was placed in a container of conically wound tungsten wire. Before vaporizing the barium, the container with its contents was thoroughly outgassed.

The barium films investigated were identical as regards the manner in which they were deposited with those on which we carried out experiments with CO to determine the influence of wall temperature. We shall discuss these experiments presently.

Influence of the getter holder; gettering-rate measurements with hydrogen

Our view that the properties of a barium getter film are largely established when the barium is still close to the getter holder, is based on gettering-rate measurements with hydrogen.

From the first series of measurements it was found that the rate with which a barium film absorbs hydrogen, as long as the film is not yet saturated, decreases linearly with the quantity of the gas already absorbed. By drawing a straight line through the values found, and extrapolating, we find the value of the initial gettering rate a_0 from the point of intersection with the gettering-rate axis. The point of intersection with the other axis gives the capacity of the film. The results of such a series of measurements, plotted in this manner, are shown in *fig. 9*. Variation of the amount of gas admitted produced no significant change in the value found for a_0, which indicates that, for the absorption of hydrogen, a_0 is constant in a fairly wide pressure range (the initial pressure varied from 10 to 100 μ Hg).

The value found for the initial gettering rate per mg barium was $(2.5 \pm 0.2) \times 10^{-2}$ l/sec.mg. (The quantity of barium in the films varied from 6.3 to 7.9 mg. The area of the films was in every case 35 cm², which means that if the barium were to form a solid, smooth layer, the thickness would be between 0.49 and 0.61 μ.)

In the second series of measurements the inside of the bulb was given a prior coating of colloidal graphite, producing a rough surface. If the barium film were a smooth homogeneous layer, the area of the barium film exposed to the gas, and hence the initial gettering rate, would now necessarily be larger than if the film were deposited on a smooth glass wall. It was found, however, that under these conditions the initial gettering rate per mg was the same as in the first series of measurements [7]).

In the third series of tests the getter holder was mounted in a large glass sphere, so that the film deposited had a substantially larger macroscopic area (186 cm²) than in the first two series. The thickness was, of course, proportionally smaller and

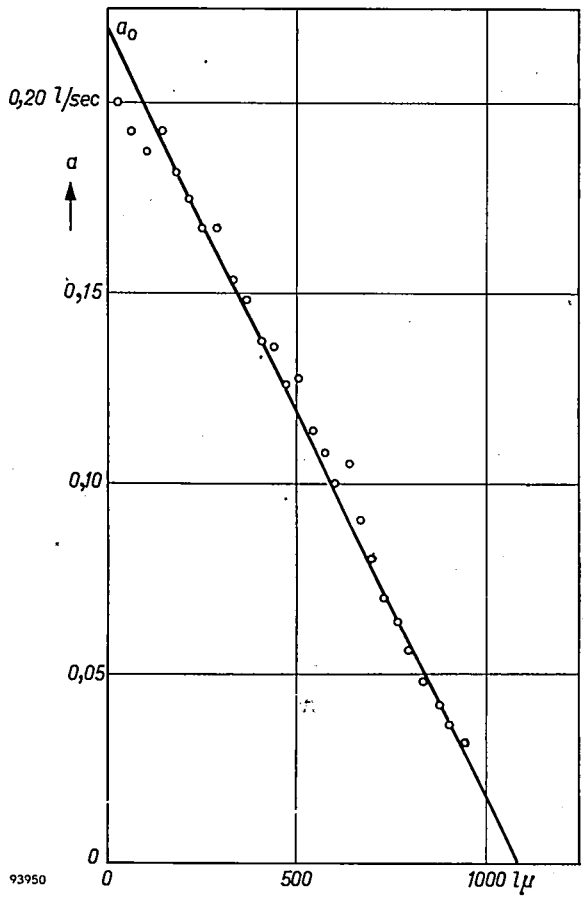


Fig. 9. The absorption of hydrogen by barium shows the appearance of a volume effect: the absorption rate (ordinate) depends linearly on the amount of gas already absorbed (abscissa). Extrapolation of the line drawn through the measured points gives the initial rate (point of intersection with vertical axis) and the total absorption capacity (point of intersection with horizontal axis).

was approximately constant over the whole surface. Here again the same value was found for the initial gettering rate per mg.

It may be concluded, then, that in all the three cases mentioned the area of the film accessible to gas (the "microscopic area") was equally large, which is in agreement with our picture of the film as a porous structure of stacked barium spheres: for, the number and magnitude of the spheres may be assumed to be equal in all three cases.

[6]) See G. Hass and N. W. Scott, J. Opt. Soc. Amer. **39**, 179, 1949.

[7]) J. Morrison and R. B. Zetterstrom, J. appl. Phys. **26**, 437, 1955, carried out an analogous experiment with MgO instead of graphite. As in our case, they found for CO no greater values for a_0 nor for the absorption capacity.

In addition it was found that the gettering rate is proportional to the total quantity of the barium not yet used up; a linear relationship exists between a and the quantity of gas already absorbed. The absorption of hydrogen thus shows the appearance of a volume effect [3]).

In the fourth series of measurements, using the same large bulb mentioned above, a combination was introduced of two or three of the strip getters used in the first three series. It was found in this case that a_0 itself as well as the initial gettering rate, related to one cm² or to one mg, had both decreased. With two strip getters the latter had dropped to about $^1/_5$, and with three strip getters to as little as about $^1/_{25}$ of the value of the first three series; a_0 itself had dropped to about $^1/_2$ and $^1/_5$ respectively of the former values. The film thickness in the tests with three strips was the same as in the first series; in the tests with two strips it was slightly less than half.

The fact that, in spite of variation in the film thickness, a constant value was found for the initial rate per mg in the first three series, whereas a different value was found when using different getter holders (in the fourth series), demonstrates that the properties of a getter film are partly determined beforehand in the immediate neighbourhood of the getter holder, and depend upon the type of holder used.

Experimental details. The barium getter films investigated in the above four series were obtained from samples of the Philips strip getter type R1 679 36, illustrated in fig. 2. To minimize variations in properties, the getters were taken from a single production run. (This was also done for all subsequent experiments.) High-frequency outgassing was not applied, but for all tests the bulb and getter holder were heated at 400 °C as soon as high vacuum was reached. The barium getter film was allowed to form as soon as the bulb had cooled to room temperature.

During the formation of the film the temperature of the bulb rose slightly owing to the barium cooling on it and to the heat radiated by the glowing getter holder. The differences in this respect between the small and large bulbs did not appear to be significant. The gettering-rate measurements were always carried out after the bulb had dropped to room temperature again. *Fig. 10a* shows the getter holders used in the fourth series of tests.

The dimensions of the small bulb are indicated in fig. 6, those of the large bulb in fig. 10*b*. In the large bulb, the getter holders were so positioned as to produce getter films of as constant a thickness as possible and with well-defined edges.

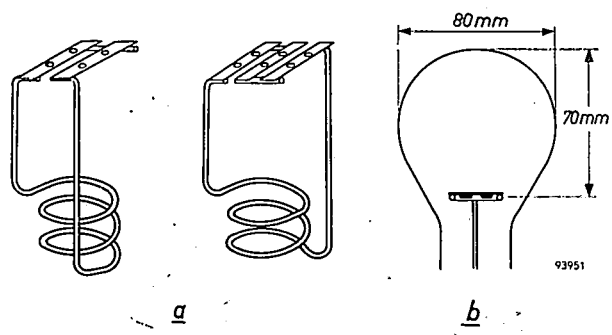


Fig. 10. *a*) Getter holders made by combining two and three strips, respectively, and provided with a coiled support for induction heating.
b) Sketch showing dimensions of the large bulb used for making a getter film of large macroscopic area and constant thickness.

Influence of the wall temperature; measurements of absorption capacity for CO

The size of the microscopic surface of barium getter films can be comparatively investigated not only by measurements of the gettering rate, as described above, but also — and more simply — by measurements of the *absorption capacity*, provided a gas is used which is known with certainty to be absorbed by the *surface*. Such a gas is carbon monoxide [8]).

According to our hypothesis regarding the structure of the barium films and the manner in which they are formed, we may expect that a film of this kind will show the following properties. In the first place the absorption capacity for CO will decrease according as the temperature of the bulb wall was higher during the deposition of the film. In the second place the absorption of a film which is so sparse that the barium grains lie separate from each other (i.e. cannot fuse together) will be large and will not be greatly affected by the temperature of the bulb during the formation of the film.

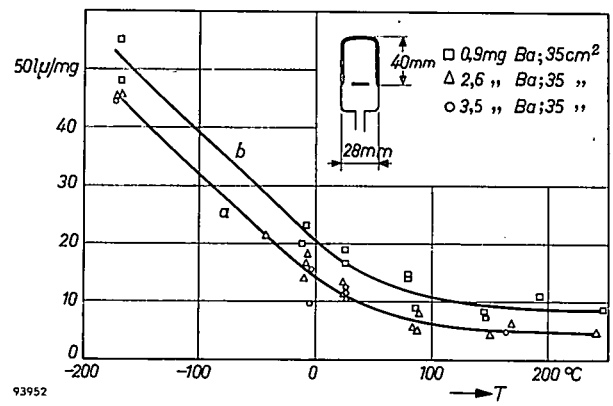


Fig. 11. Measurements of absorption capacity for carbon monoxide. The absorption capacity per mg barium, measured at room temperature, is plotted against the temperature of the bulb during the formation of the barium film. In all cases the getter holder was 40 mm away from the end of the tube. The values found for films containing 3.5 and 2.6 mg barium show no significant difference (curve *a*). Curve *b* applies to films containing only 0.9 mg barium.

The first-mentioned effect is clearly demonstrated in curve *a* of *fig. 11*. In this figure the absorption capacity per mg — all measurements, as with hydrogen, at room temperature — is plotted against

[8]) See S. Wagener, J. phys. Chem. **60**, 567, 1956.

the temperature of the bulb wall for two kinds of film (containing respectively 3.5 and 2.6 mg barium and originating from two types of getter holders). It is striking that the values found with both kinds of film coincide within the accuracy of measurement and are higher the lower was the temperature.

The properties of films so sparse that the barium grains lie separate from each other (obtained from a third type of getter holder), appear from the measurements represented by curve *b* in fig. 11 and by *fig. 12*. The measurements first mentioned were made on films which were transparent over a

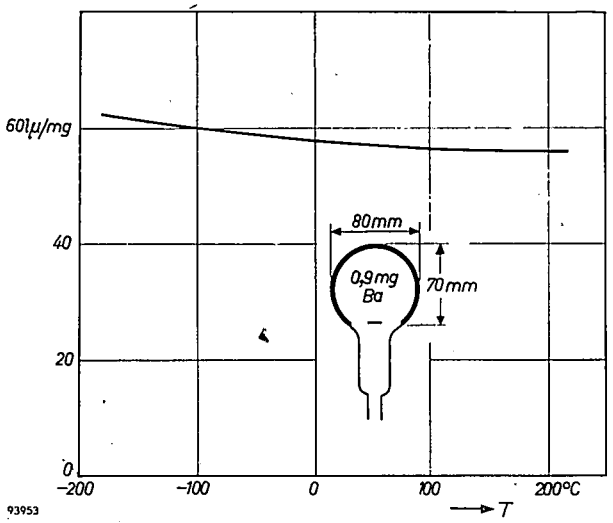


Fig. 12. Absorption capacity per mg of barium films whose grains are not in contact with each other. It can be seen that in this case the temperature of the bulb wall during the formation of the film has only a negligible effect on the absorption capacity.

part of their surface (0.9 mg Ba on 35 cm²), which was not the case with the films represented by curve *a*. The variation of absorption capacity with temperature is completely identical with that of curve *a*, but all values are somewhat higher; this is understandable if we assume that the barium grains lie free in the transparent part of the film.

The measurements recorded in fig. 12 were made on extremely thin films (0.9 mg Ba on 262 cm²). We see that in fact the absorption capacity per mg is again larger than found by the measurement recorded by curve *b* in fig. 11. Furthermore the absorption capacity, as expected, is here almost independent of the temperature of the bulb wall during the formation of the film.

Experimental details. The above experiments with CO, and some others still to be discussed, were carried out with getter holders of three different types (Philips R1 679 50, R1 679 51 and R1 679 82). They were heated for a time T of 11 sec, the "vaporizing time" t being 7 sec [9]). A number of measurements showed that the quantities of barium vaporized under these conditions were respectively 3.5 ± 0.2 mg, 2.6 ± 0.1 mg and 0.9 ± 0.1 mg.

The set-up for these measurements (see *fig. 13*) differed slightly from that used for the hydrogen experiments. The quantity of gas (coming from the reservoir B via the lock L formed by the stopcocks g_1 and g_2, and further via the three-way cock g_4 and the liquid-air traps R_1 and R_2) which was admitted to the barium film G, was in this case determined beforehand by a pressure measurement with a MacLeod gauge M. The extra trap R_2 was put in operation after outgassing of space C (heating to 400 °C) and before the depositing of the barium film. This was always done in those experiments whereby the space C was at a temperature below room temperature, and frequently also in the other cases for test purposes.

In the experiments on the influence of the bulb temperature the distance from the getter holders to the apex of the bulb was invariably 40 mm, and the area covered by barium was in all cases 35 cm².

Further experiments with CO

It has long been known that very good getter films are obtained when the valve in which the getter is vaporized is filled with an inert gas under suitable pressure [10]).

In order to compare the quality of these films quantitatively with that of others, we performed a number of experiments the conditions of which, apart from the gas filling of the valve, were identical with the experiments discussed above on the influence of the bulb temperature. The valve was filled with argon under a pressure of 6 mm Hg.

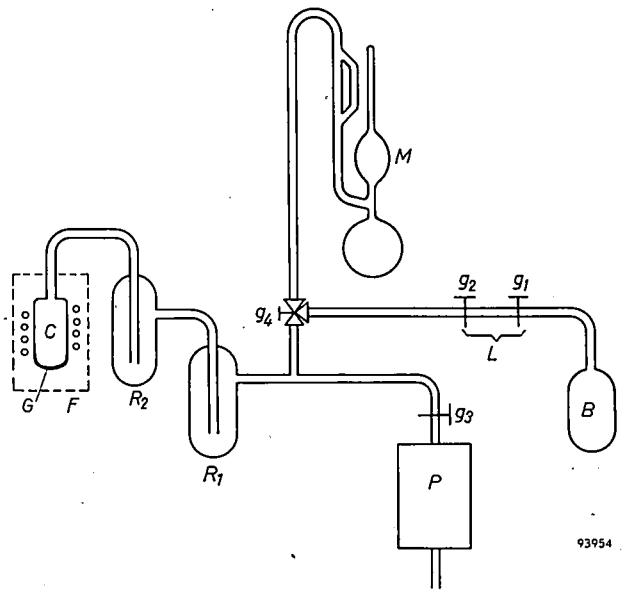


Fig. 13. Set-up used for measuring the absorption capacity for carbon monoxide. The letters have the same meaning as in fig. 6. The amount of gas admitted to C can be accurately controlled by the three-way cock g_4 and the MacLeod gauge M. The extra liquid-air trap R_2 was found necessary in those experiments in which the bulb wall was at a temperature lower than room temperature. F represents the heating or cooling device, as the case may be.

[9]) See page 292.

[10]) A. L. Reimann, Phil. Mag. **18**, 1117, 1934. See also the article by De Graaf and Hamaker, quoted under [4]).

The absorption capacity per mg was found to be not much larger than that found in the earlier experiments with very thin films.

This leads to the conclusion that the use of this gas filling does not essentially improve the properties of a barium getter film; on the contrary, under normal circumstances the best results are frequently not obtained, that is to say the mist of barium droplets is not deposited on the bulb in the most effective way.

Vaporization in a gas atmosphere gives rise, as a rule, to a black getter film. If the gas pressure is relatively low, a "normal" film is produced, which however also exhibits a large absorption capacity for CO. It is evident from the above, therefore, that there is no justification in regarding black getter films as belonging to a distinct class as regards their microscopic area. As mentioned earlier, the getter films of commercial electronic valves may be regarded effectively as vaporized through gas.

On the basis of our hypothesis on the structure and manner of formation of a barium getter film, which is confirmed by the experiments described, it is possible to explain the results of the following more complicated experiments. First of all a series of experiments was carried out in which the distance from the getter holder to the bulb wall was varied. Two distances were adopted, the getter holder being mounted 6 and 40 mm away from the apex of the bulb. The getter holders used in these experiments were the types earlier mentioned (yielding respectively 3.5 mg, 2.6 mg and 0.9 mg barium), so that we were able to make six types of getter film. During the evaporation of the barium the tube was neither heated nor cooled. Four of the six kinds of film were found to have almost the same absorption capacity for carbon monoxide per mg barium (12 l.μ/mg). The 3.5 mg film at a wall-to-holder distance of 6 mm was poorer (8 l.μ/mg), the 0.9 mg film at a wall-to-holder distance of 40 mm was better (18 l.μ/mg). In the first of these two cases a great deal of barium was present per cm^2 with the intensively radiating getter holder at a short distance from the bulb wall. In all probability this caused, during the vaporization, greater heating of the bulb and of the barium already deposited than in the other experiments. As we found earlier, this indeed results in a reduced absorption capacity. In the second case just the opposite was the case.

A further noteworthy result is that the same absorption capacity per mg is found for the 2.6 mg films at both wall-to-holder distances, although the films differ by a factor of 4 in surface area. This accords entirely with what we found earlier with regard to the absorption rate for hydrogen when the barium was vaporized in a small and in a large bulb. This result adds support to the view that measurements both of the gettering rate for hydrogen and of the absorption capacity for carbon monoxide can give a (relative) measure of the size of the area accessible to gas. The size of this "microscopic" area is evidently substantially independent of the macroscopic area of the film.

The following series of experiments which we shall discuss to conclude this section are a repetition of the earlier experiments on the influence of the bulb temperature, but now made with the getter holders (again the same three types) at a distance of only 6 mm (instead of 40 mm) from the bulb wall. The results are set out in *fig. 14*, with curve *a* from fig. 11 included for comparison. It can be seen that, as opposed to the curves in fig. 11, the absorption capacity per mg in this case tends to a maximum at wall temperatures below about —80 °C.

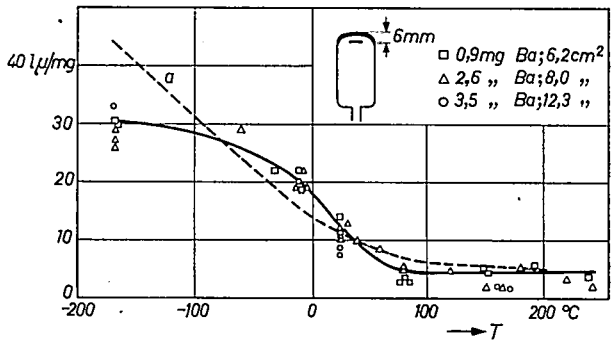
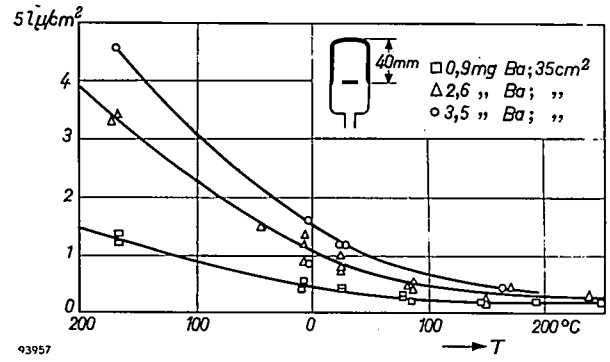


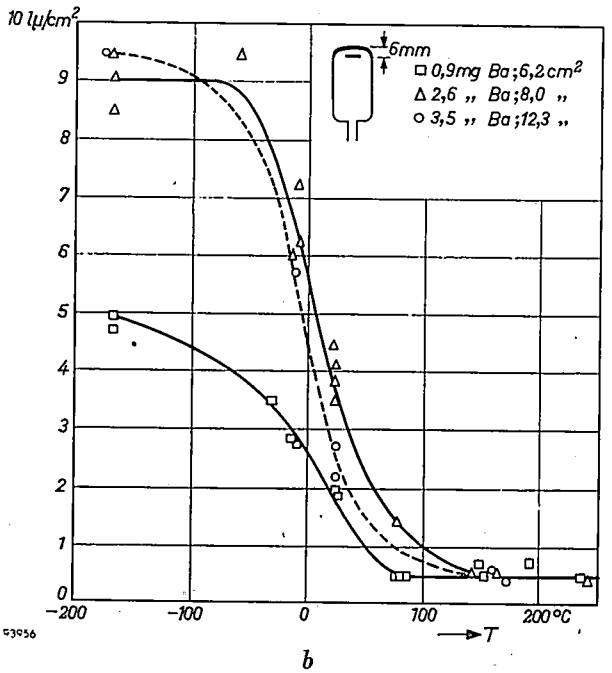
Fig. 14. Absorption capacity per mg of thick barium films as a function of the wall temperature during formation of the film. As in fig. 11, the plotted points refer to films containing 3.5 mg, 2.6 mg and 0.9 mg barium. The full curve is drawn to fit the points for the 0.9 mg films, but also applies fairly well to the others. Curve *a* from fig. 11 (dashed) is included for comparison.

This is due in all probability to the intense heat radiation from the getter holder. The fact that the values found at lower wall temperatures are also appreciably lower than in the earlier experiments points in the same direction. It is remarkable that between —80 °C and room temperature the absorption capacity per mg is higher than that of the thinner films of fig. 11. This is presumably caused by the action of the gas escaping from the getter holder during vaporization, because curves relating to later experiments with getters largely outgassed beforehand do not show this "hump". It appears from this that the position of a getter holder which has not been outgassed affects the quality of the film in an even more complicated way than could be deduced from the earlier experiments; the action of the gases released from the getter holder is apparently not exclusively determined by the method of heating and the nature and quantity of these gases.

If the absorption capacities measured in these experiments be considered not with respect to the quantity of barium but with respect to the macroscopic size of the film surface, it is found that at high temperatures the absorption capacities per cm^2 of all three kinds of film approach one and the same limiting value (approximately 0.5 $l.\mu/cm^2$). Remarkably enough, this value is higher than that found in the experiments of fig. 11 which was about 0.3 $l.\mu/cm^2$ (getter holders at a distance of 40 mm from the bulb wall); see *fig. 15*. The reason for this may be the thin spreading-out of the film edges: in the experiments to which fig. 11 refers, the edges of the films were much more sharply defined and therefore less troublesome. In view of the dimensions involved, this thin spreading-out of the films might well be a consequence of the escape of gases during the vaporization of the barium.



a



b

Fig. 15. *a*) Absorption capacity per cm^2 (macroscopic area) of the films for which the capacity per mg is given in fig. 11. *b*) Idem for the films in fig. 14.

Reaction of barium with carbon monoxide; calculation of grain diameter

According to Morrison and Zetterstrom [11]), barium may react with carbon monoxide in one of the following three ways:

$$\begin{aligned} Ba + CO &\rightarrow BaO + C, \\ 3\,Ba + 2\,CO &\rightarrow BaC_2 + 2\,BaO, \\ Ba + 2\,CO &\rightarrow Ba(CO)_2. \end{aligned}$$

From the results of some experiments now to be described it is our conclusion that it is the second of these reactions that actually takes place.

A barium film was deposited (at room temperature) in a bulb which had previously been heated to 400 °C. After raising the film to a temperature of 320 °C, a quantity of CO was admitted into the bulb. As soon as the film was saturated [12]), the CO that had not been absorbed was removed and water vapour admitted. Examination of the bulb contents with a mass spectrometer revealed that a chemical reaction had taken place: apart from water vapour, C_2H_2 was present. This gas could not have arisen if the barium and carbon monoxide had reacted according to the first or the third of the above three equations. It can, however, arise from the BaC_2 formed in the second reaction, according to the equation:

$$BaC_2 + H_2O \rightarrow BaO + C_2H_2\uparrow.$$

If we calculate how much CO can be absorbed by one mg barium in the second of the above three equations, we find 81.9 lμ. The result obtained from a number of experiments was 81 ± 7 lμ. This, too, indicates that the absorption takes place according to the second reaction.

Naturally this is still no proof that the said reaction is also responsible for the CO absorption at room temperature. We have reason to assume, however, that this is in fact the case: in the first place because of the fact that films with a large microscopic area exhibit the same change of appearance in the case of absorption of CO at room temperature (they become transparent with a light brown colouring), and secondly because of provisional results of a mass-spectrometer investigation now under way [13]).

---

[11]) See page 440 of the article quoted under [7]).

[12]) The temperature of 320 °C is high enough to ensure the reaction of all the barium and at the same time far enough below the temperature at which the bulb was outgassed (400 °C) to prevent the release of unwanted gases from the bulb wall.

[13]) These results will be published in due course elsewhere. Bloomer, Brit. J. appl. Phys. **8**, 352, 1957 (No. 9), believes that the reaction takes place according to the first of the above three equations.

It is expected that this investigation will explain, among other things, why, in our experiments, H_2 and CH_4 (and no C_2H_2) are formed when water vapour is admitted to barium getter films saturated with CO at room temperature. Such reactions with water are a possible reason for the fact that hydrocarbons are sometimes found in the residual gas (the gas not absorbed by the getter film) in radio valves.

A barium getter film "saturated" with CO at room temperature is also unable to absorb any further hydrogen or nitrogen, unless the temperature is raised. In this case an occlusive layer has apparently formed on the surface. Starting from the second reaction equation mentioned above and the lattice constant of barium (5×10^{-8} cm), we can calculate the thickness of this layer if we are able to measure the absorption capacity of a film whose active surface area is equal to the macroscopic area. This is the case with the films formed at a high bulb temperature. The limit value found here was approx. 0.3 l.μ/cm^2 (cf. fig. 15). It follows, then, that the layer in question contains about 36 atomic planes, i.e. its thickness is about 18 lattice spacings $= 9 \times 10^{-3}$ μ [14]). (The barium lattice is body-centred cubic.)

Diameter of the barium grains

If we assume that the result just found is also applicable to the grains from which we imagine a barium getter film to be composed (the grains not being fused together), we can calculate the diameter of these grains in the following way.

Let D be the average diameter of the grains, and d the diameter of that interior part of the grain which is inaccessible once the layer has formed, then the non-reacting fraction of the quantity of barium is $\frac{1}{6}\pi d^3/\frac{1}{6}\pi D^3$. On the other hand, from the measurements of the absorption capacity at 320 °C (about 81 l.μ/mg) and at room temperature (approximately 57 l.μ/mg for the extremely thin film mentioned earlier) we know that this fraction is about 24/81. A second relation between d and D follows from the known thickness of the layer: $D - d = 2 \times 9 \times 10^{-3}$ μ. The value of the grain diameter D found by solving these two equations is roughly 0.05 μ. If we take for the absorption capacity at room temperature the value found for the films formed in an argon-filled bulb (approx. 70 l.μ/mg), we find D to be about 0.04 μ. Both values are in good agreement with those which we derived from the photomicrographs obtained with the electron microscope. Assuming the diameter to be 0.05 μ, we then find that one mg barium is divided into 4.13×10^{12} grains. These have a total (microscopic) area of 325 cm^2 and would cover, if laid side by side in one layer, a wall area of 89 cm^2.

When applying in practice the results reported in this article it is necessary first of all to ascertain what gases are to be absorbed by the barium getter film. The absorption capacity for hydrogen and oxygen depends on the total quantity of barium, that for carbon monoxide on the size of the microscopic area. We have found that circumstances can arise in which a larger microscopic area is obtained when less barium is vaporized.

It has been found that better results are obtained as regards gettering rate (and also as regards the absorption capacity for CO) when the barium is vaporized after the tube has been sealed off and is at room temperature, than when the tube is still on the vacuum pump. This may also be explained in the light of the results given above.

Summary. The properties of the barium film deposited in vacuum tubes in order to absorb gases released after the tubes have been sealed off, depend in a complicated way on the manner in which the film is formed. The influence of the method of formation can be understood by assuming that the film consists of a stacked structure of barium grains partially fused together. It is suggested that these minute grains are solidified droplets of barium which are created not on the surface on which the film is deposited but in the immediate vicinity of the getter holder upon vaporization of the barium. The authors attribute this to the presence of gases escaping from the getter holder together with the barium. Experiments concerning the gettering rate for H_2 under various conditions all point in this direction.

The colder the wall on which these droplets are deposited, the less will they fuse together and the larger will be the "microscopic" area of the barium film. This is confirmed by experiments on the absorption capacity at room temperature for CO (a gas which is absorbed only at the surface). Films that are so thin that the spheres are not in contact with each other — and hence cannot fuse together — show a high absorption capacity for CO. As expected, the absorption capacity in this case is but little dependent on the temperature of the bulb wall during vaporization.

Films formed in valves containing argon (under 6 mm Hg pressure) have a high absorption capacity, although it is not much higher than that of the best "normal" getter films. It is concluded from this that vaporization in argon does not essentially improve the quality of a getter film, but rather that with "normal" vaporizations the best results are not as a rule obtained, i.e. the barium mist is not deposited on the wall in the most effective manner. Photomicrographs of an SiO replica of the surface of barium films, made with an electron microscope, show that the surface is indeed composed of small grains, which have a diameter of about 0.04 micron.

The absorption of CO takes place in all probability according to the reaction $3\,Ba + 2\,CO \rightarrow BaC_2 + 2\,BaO$. On this basis it can be calculated that the barium layer so transformed is about 18 lattice spacings thick, i.e. 9×10^{-3} μ. From this thickness, from the absorption capacity of a good barium film at room temperature, and from the absorption capacity of the same quantity of barium at 320 °C (at which temperature all the barium reacts), the grain diameter is calculated to be 0.04 to 0.05 μ.

In order to describe the properties of a barium getter film in more detail than hitherto, the authors introduce, in addition to the quantities gettering rate and absorption capacity, corresponding specific quantities with respect to both one mg barium and one cm^2 of the barium-covered surface.

The quantities vaporization time and heating time are introduced to characterize the way in which the getter holders are heated to vaporize the barium; the "melt" curve indicates the moment at which heating must be stopped to prevent the getter holder from melting.

[14]) Wagener, J. phys. Chem. **60**, 567, 1956, finds 27 atomic layers, i.e. 13½ lattice spacings.

VOLUME 19, 1957/58, No. 11 pp. 301-332 Published 31st May 1958

Philips Technical Review

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

PARAMAGNETIC RESONANCE

by J. S. van WIERINGEN.

538.222

The investigation of the phenomenon of paramagnetic resonance, discovered in 1945, has proved to be a valuable tool in the study of the solid state. For this phenomenon to occur the substance must contain a number of "atomic magnets" (paramagnetic atoms or ions, colour centres, donors or acceptors in semi-conductors, organic radicals). The specimen for examination is placed in a microwave resonant cavity, to which energy is supplied via a waveguide, the whole being placed in a magnetic field of variable strength. The resonance spectrum (reflected power as a function of the applied field) then provides important information on the properties of the crystal lattice.

The phenomenon of paramagnetic resonance

In 1945 Zavoisky discovered the phenomenon, predicted by Gorter and Kronig as early as 1936 [1]), that paramagnetic ions in a crystal subjected to a constant magnetic field (of induction B) are capable of absorbing energy from a superimposed high-frequency alternating magnetic field (of frequency f_r). This occurs when the following relation is fulfilled:

$$f_r = g\frac{e}{4\pi m}B, \quad \ldots\ldots \quad (1)$$

where e is the charge on the electron (1.6×10^{-19} coulomb), m its mass (9×10^{-31} kg), and g a factor approximately equal to 2. Since this absorption occurs in a very narrow frequency range, it is referred to as paramagnetic resonance absorption, or simply as paramagnetic resonance. It was later found that the phenomenon occurs not only with paramagnetic ions, but also with colour centres, with donors and acceptors in semiconductors, and with organic radicals — all being cases in which "atomic magnets" are present in the substance. With an applied field $B \approx 0.3$ Wb/m^2 [3000 gauss], the frequencies f_r are of the order of 10^4 Mc/s.

In a given magnetic field the absorption usually takes place not at one specific frequency given by (1), but rather at a number of frequencies f_1, f_2 etc. in the neighbourhood of f_r. The paramagnetic absorption line is thus split into several components; the resulting pattern, defined by the differences f_1-f_r, f_2-f_r, etc. depends on various factors, in the first place on the nature of the atomic magnets, and further on the crystal lattice in which the atomic magnets are situated. Moreover, the line width of the components also varies, this depending on the crystal lattice, the concentration of the atomic magnets, lattice defects and other factors. A study of the paramagnetic absorption spectra, therefore, enables one to obtain information on the electrical fields in which the ions in the crystal are situated, and on the behaviour of the electron orbits in these fields. The investigation of paramagnetic resonance is thus an additional method for the study of the solid state, and takes its place alongside other methods such as the investigation of electrical conductivity, thermo-electric power, Hall effect, X-ray diffraction and so on.

In view of the increasing interest in the properties of solids it is not surprising that a vast amount of research has been carried out in the field of paramagnetic resonance since 1945. More than 400

[1]) C. J. Gorter and R. de L. Kronig, Physica **3**, 1009, 1936; E. Zavoisky, J. Phys. USSR **9**, 211, 1945 (also ibid. **10**, 170 and 197, 1946).

publications have appeared on the subject [2]).

In this article we shall discuss some aspects of this research, including investigations carried out in the Philips laboratories at Eindhoven. First of all we shall briefly review the theory of paramagnetic resonance.

For the readers' convenience some facts on paramagnetism and ferromagnetism are recapitulated here.

Besides an electric charge an electron possesses a magnetic moment, associated with an angular momentum (spin). Since atoms consist of a nucleus, surrounded by an electron cloud, it may be expected that all atoms will exhibit magnetic properties, not only because of the magnetic moments of the individual electrons but also because the electrons describe orbits around the nucleus, which we may regard as circular currents.

In general, free atoms are found to have a magnetic moment which is the resultant of the spin and orbital moments of the individual electrons. It should be recalled here that most of the electrons move in closed shells and that the resultant of the spin and orbital moments for such a shell is zero. For instance, the atoms of the inert gases have no resultant moment since their electron cloud consists entirely of closed shells. The same applies to the ions of which the crystal lattices of most inorganic solids are built up. There are, however, ions whose electron cloud does not have the "inert-gas structure", one or more of their shells not being completely filled. These ions show a resultant moment which can be oriented in an external magnetic field. Substances containing these ions thus have a positive magnetic susceptibility, i.e. they are paramagnetic. The most familiar paramagnetic ions are those in compounds of the iron group (e.g. V^{3+}, Cr^{3+}, Mn^{2+}, Fe^{2+}, Fe^{3+}, Co^{2+}, Ni^{2+}), in which the outermost electron shell is incompletely occupied, and the rare-earth ions, which owe their paramagnetic properties to the fact that one of the inner shells is not completely filled.

Some metals are also paramagnetic. Generally speaking, metals consist of an ion lattice between which electrons move more or less freely. The orbital motion of most of the electrons is such that their spin and orbital moments compensate each other — analogous to the compensation for the majority of electrons in the electron cloud of a free atom — and only a few electrons are able (particularly at higher temperatures) to contribute to the paramagnetism. As a rule, therefore, the paramagnetism is weak (Na, Ca) [3]).

In this article we are not concerned with the paramagnetism of metals, but rather with that of certain semi-conductors, which is due to the presence of foreign atoms (donors, acceptors) or to the electrons or "electron holes" ceded to the lattice by these atoms; these electrons or holes function as atomic magnets.

Where the interaction between the atomic magnets in a substance is such that they are in spontaneous parallel orientation, the substance is ferromagnetic. In that case, even in the absence of an external magnetic field, a macroscopic magnetic moment is present in the Weiss domains into which the substance is divided. Ferromagnetism, which occurs only in a few metals (Fe, Co, Ni, Gd) and certain compounds thereof, in certain Mn alloys and in ferrites, will be touched on in passing in the following discussion of the theory.

Theory of paramagnetic resonance

In general, paramagnetism is due to the spin and orbital moments of the electrons. For reasons which will later be discussed, the orbital moment can usually be neglected to a first approximation in solids.

According to quantum mechanics, the angular momentum of one ion (in our case the resultant of the electron spins of that ion) can only assume discrete values. Quite generally, the angular momentum may be represented by $\sqrt{S(S+1)}\,(h/2\pi)$, where S is a multiple of $\frac{1}{2}$.

Let us first consider a free electron, in which case $S = \frac{1}{2}$. In a magnetic field B an electron can take up only two kinds of positions, the component of the angular momentum in the direction of B being either $\frac{1}{2}(h/2\pi)$ or $-\frac{1}{2}(h/2\pi)$. The component of the magnetic moment is proportional to this angular momentum, the constant of proportionality being $g_e(e/2m)$, where $g_e = 2.0023$. The two positions S correspond to two energy levels:

$$E = \pm g_e \frac{e}{2m} \times \tfrac{1}{2} \frac{h}{2\pi} B. \quad . \; . \; . \quad (2)$$

Hence the difference in energy between the two levels is

$$\Delta E = g_e \frac{eh}{4\pi m} B. \quad . \; . \; . \; . \; . \quad (3)$$

Similar reasoning applies to an arbitrary paramagnetic ion. This can take up a number of positions with respect to the field B, such that the component of the angular momentum in the direction of the field is equal to $m_s(h/2\pi)$. The number m_s can assume values between $+S$ and $-S$, such that the difference between successive values is equal to 1. For example, where $S = \frac{5}{2}$, m_s may be equal to $\frac{5}{2}$, $\frac{3}{2}$, $\frac{1}{2}$, $-\frac{1}{2}$, $-\frac{3}{2}$ or $-\frac{5}{2}$, totalling $2S+1 = 6$ values. The energy difference between successive levels is again given by (3). *Fig. 1* shows schematically the possible orientations of the angular momentum with respect to the field, and the associated energy levels for a number of values of S, i.e. for a number of different ions. (Different values of S can also be found for one and the same ion, depending on the electron distribution; in general, an atom in an excited state

[2]) We cite here only a number of survey articles covering the subject:
B. Bleaney and K. W. H. Stevens, Paramagnetic resonance, I, Rep. Progr. Phys. **16**, 108-159, 1953.
K. D. Bowers and J. Owen, Paramagnetic resonance, II, Rep. Progr. Phys. **18**, 304-373, 1955.
J. E. Wertz, Nuclear and electronic spin magnetic resonance, Chem. Rev. **55**, 829-955, 1955.
D. J. E. Ingram, Spectroscopy at radio and microwave frequencies, Butterworth, London 1955.

[3]) In some metals (e.g. Cu) the weak paramagnetism is dominated by the diamagnetism present in all substances; this is due to the circumstance that an external magnetic field gives rise to a weak and oppositely oriented magnetic moment in the electron cloud of the atom (negative susceptibility).

may have a value of S different from that obtaining for the ground state.)

According to quantum theory, the emission of electromagnetic radiation takes place during the transition from one state with energy E_2 to as tate with energy E_1 ($E_2 > E_1$), and absorption takes place with the reverse transition. The frequency f of the emitted or absorbed radiation is determined by the relation [4])

$$hf = E_2 - E_1. \quad \ldots\ldots \quad (4)$$

Now a quantum-mechanical treatment shows that, in each case, transitions are possible only between successive energy levels. The frequencies are therefore always given by $hf = \Delta E$, where ΔE is given by (3).

Transitions between the levels can occur when an alternating magnetic field is applied at right angles to the field B, the frequency f_r of the alternating field being given by $hf_r = \Delta E$, i.e.,

$$hf_r = g_e \frac{eh}{4\pi m} B. \quad \ldots\ldots \quad (5)$$

During these transitions, energy in the form of quanta hf_r is absorbed from the alternating field or given up to it. This energy-absorption phenomenon is essentially the paramagnetic resonance absorption. In fact, the condition (5) is identical with (1) if we substitute for g the accurate value g_e (= 2.0023).

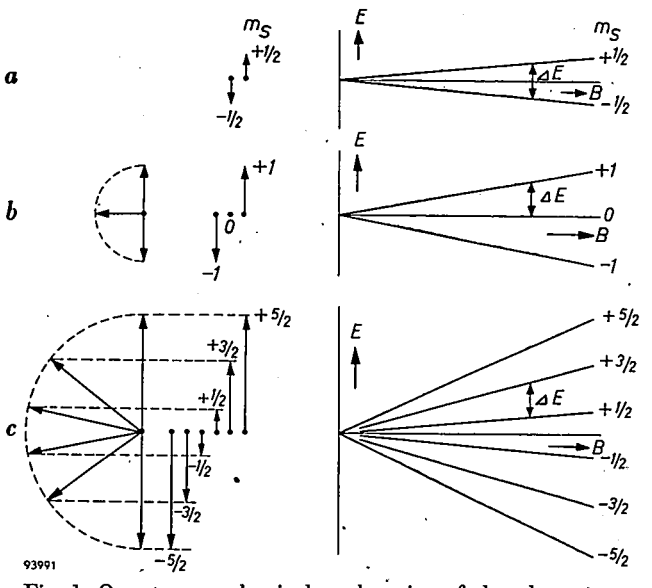


Fig. 1. Quantum-mechanical explanation of the phenomenon of resonance. A dipole having an angular momentum (quantum number S) and an associated magnetic moment can orient itself in only $2S+1$ different ways with respect to an external magnetic field B. The component of the angular momentum parallel to B, $m_s(h/2\pi)$, can assume any of $2S+1$ discrete values, where $S > m_s > -S$. This gives rise to $2S+1$ energy levels, and transitions are possible only between successive levels, of energy difference ΔE. The resonance frequency is given by $hf_r = \Delta E$. The possible orientations of the dipole and the associated energy levels as a function of B are drawn for *a*) $S = \frac{1}{2}$ (case of a single electron), *b*) $S = 1$, *c*) $S = \frac{5}{2}$ (case of the Mn^{2+} ion).

In order to be able to observe resonance absorption an additional effect must appear, for, at resonance, the paramagnetic ions are also continually yielding energy quanta hf_r to the field (so-called stimulated emission [5])). Absorption can only be observed when the number of absorbed quanta exceeds that of the emitted quanta so that, on the average, energy is absorbed from the field. This can indeed occur by interaction between the atomic magnets and their environment.

Let us first consider N entirely free electrons, so that we have two energy levels, differing by ΔE as given by (3), the lower level being occupied by n_1 and the upper by n_2 electrons ($n_1 + n_2 = N$). A high-frequency magnetic field with frequency f_r causes energy transitions in both directions with equal probability. If the lower level was initially more densely occupied ($n_1 > n_2$), then n_1 will decrease and n_2 increase, until $n_1 = n_2 = \frac{1}{2}N$. The number of absorption processes, which is determined by n_1, is then equal to the number of emission processes, which is determined by n_2, and the net result is that no energy is absorbed from the field.

In the case of ions forming part of a crystal lattice, interaction will occur between the electrons and the lattice, as a result of which a state of thermal equilibrium will tend to be set up. For the sake of simplicity we assume that we are concerned with ions, the paramagnetism of which is due to only one electron ($S = \frac{1}{2}$). Hence there are again two energy states. In thermal equilibrium,

$$n_2/n_1 = \exp(-\Delta E/kT), \quad \ldots\ldots \quad (6)$$

where k is Boltzmann's constant ($k = 1.38 \times 10^{-23}$ J/°K).

In this case, according to (6), $n_1 \gg n_2$ in the state of equilibrium. The number of absorptions therefore exceeds that of the emissions, so that now the net result is that energy is absorbed from the field. Although the predominating absorption tends to increase the occupation n_2 of the higher level, the absorption nevertheless continues, since, owing to the interaction with the crystal lattice, a portion of the absorbed energy is given up to the lattice as heat. This process (paramagnetic relaxation) is usu-

[4]) This is the familiar Bohr formula. In atomic theory the symbol ν is usually used for the frequency. Since in the present case the frequencies lie in the region of radio waves, the symbol f common in electrical engineering is used here.

[5]) Besides stimulated emission there is, in general, emission due to spontaneous transitions. This is important in optical spectra, but in our case it may be neglected, the frequencies and hence the probabilities being much lower.

ally so fast that, at the high-frequency energies commonly used, the ratio n_2/n_1 differs only very slightly from the equilibrium ratio (6), in spite of the presence of the alternating magnetic field [6]).

According to equation (5) it should be possible to observe the effect of paramagnetic resonance absorption even at relatively low frequencies and correspondingly low magnetic inductions. A limit is set to this, however, by the fact (which does not arise in the simple theory given here) that all transition probabilities diminish with decreasing frequency, so that the whole effect finally becomes too small to be detected. This is the reason that paramagnetic resonance was not discovered until sufficient advances had been made in the development of microwave techniques (frequencies of the order of 1000 Mc/s). Moreover, it is only at such very high frequencies that the apparatus has sufficient resolving power for observing the finer details of the resonance phenomenon, which will be discussed below and on which the practical application of paramagnetic resonance is based.

Relation between paramagnetic resonance and other phenomena

Zeeman effect

In 1896 the Dutch physicist Zeeman discovered that the spectral lines of substances split up into two or more components when the atoms are subjected to a magnetic field. The amount Δf of the splitting (difference between the frequency of a component and the frequency of the unsplit line) is given by the formula:

$$\Delta f = a \frac{e}{4\pi m} B, \quad \ldots\ldots\ldots \quad (7)$$

in which a is a numerical factor equal to a small integer or to a simple fraction. This phenomenon of "magnetic splitting" (Zeeman effect) was also found to occur in those crystals exhibiting sharp absorption lines.

The similarity of equations (7) and (1) points to a close relationship between the two phenomena. The nature of this relationship is discussed below.

We have seen how an energy level of the atom (or ion) in question is split up in a magnetic field into several sub-levels. This process occurs at each of the energy levels corresponding to different electron distributions (orbits) in the atom; see *fig. 2a*. With the Zeeman effect, as well as with paramagnetic resonance, transitions are observed between these sub-levels. In the case of paramagnetic resonance, however, the transition takes place between two sub-levels which have originated from *one* energy level, usually the lowest, of the atom; during the transition only the orientation of the magnetic moment changes (R in fig. 2*a*). With the Zeeman effect, on the other hand, the transitions take place between the sub-levels which have originated from *different* energy levels (e.g. from the ground state and an excited state — σ, π in fig. 2*a*); during the transition there is a change not only in the orientation of the magnetic moment but also in the electron distribution in the atom.

It can be seen that the energy difference — and hence the frequency — is much greater for the Zeeman transitions than for the paramagnetic resonance phenomenon. In the first case the frequencies come within the optical range, which is why the Zeeman effect was discovered so much earlier than paramagnetic resonance. Owing to the greater probabilities of transitions in the optical range, the Zeeman effect is a "strong" effect in that it can be observed in rarefied gases. For paramagnetic resonance to be observable, higher concentrations of atomic magnets are necessary, as in solids, for example. In that case the Zeeman effect is often difficult to observe because of the considerable broadening of the energy levels, particularly those of excited states.

In special cases paramagnetic resonance has been used to produce transitions between sub-levels originating from a higher (excited) level [7]). A particular case in point is mercury vapour. The changed occupation of the sub-levels in the excited state becomes manifest in this case in a change of the relative intensity of the Zeeman components (see fig. 2*b*). This is indeed a very elegant confirmation of the relation between the two phenomena.

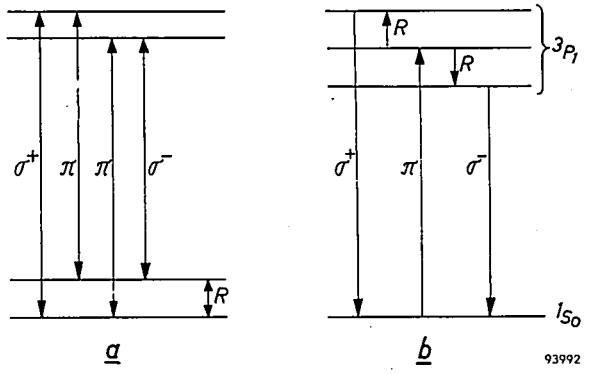


Fig. 2. *a*) Energy levels of a paramagnetic atom in a magnetic field. The ground level and one excited level are shown, both of which are split into two sub-levels by the magnetic field. The long arrows represent transitions for the Zeeman effect, and the short arrow R represents the transition between the sub-levels of the ground state, corresponding to paramagnetic resonance. The components π vibrate in the direction of the field, the components σ perpendicular thereto.
b) Experiments by Brossel and Bitter on the modification of the Zeeman pattern caused by transitions occurring between the sub-levels of an excited level in mercury vapour. Polarized radiation of 2537 Å excites mercury atoms via the π transition into the intermediate of the three $^3P_1$ levels. They drop back via the π transition. However, if the transitions R between the $^3P_1$ sub-levels are made to occur simultaneously, the two σ transitions also occur. These can be detected by the fact that their polarization differs from that of the π transition.

Ferromagnetic resonance

The phenomenon of paramagnetic resonance is also closely related to the phenomenon of ferromagnetic resonance, predicted in 1935 by Landau and Lifshitz and subsequently found with various ferromagnetic substances. Several articles on

[6]) This is opposed to the case of "nuclear spin resonance", where there are appreciable shifts in the equilibrium ratio. See H. G. Beljers, Measurement of magnetic fields by the proton resonance method, Philips tech. Rev. **15**, 55-62, 1953/54.

[7]) J. Brossel and F. Bitter, A new "double resonance" method for investigating atomic energy levels. Application to Hg $^3P_1$, Phys. Rev. **86**, 308-316, 1952.

ferromagnetic resonance (also referred to as gyromagnetic resonance) have appeared in this Review [8]).

Since this phenomenon is concerned with the magnetic moment of all the electrons in a macroscopic domain (Weiss domain), it can be explained in terms of classical as well as quantum theory. The total angular momentum, with the resultant magnetic moment parallel to it, then behaves like a spinning top whose axis, under the influence of a constant external magnetic field, describes a cone with a precession frequency

$$f_p = g_e \frac{e}{4\pi m} B. \quad \ldots\ldots\ldots \quad (8)$$

It is clear that by means of a field rotating in unison with the precession, or by an alternating field at right angles to B, it is possible to influence this motion, and that mechanical resonance (now in the literal meaning of the term) will occur if the frequency f_r of the alternating field is equal to f_p; we then have equation (1) once more.

With a certain reservation, this classical treatment can also be applied to the atomic magnets discussed above, which makes the term paramagnetic "resonance" rather more acceptable. We have preferred to give the quantum-mechanical description, however, since it is indispensable for studying the details of paramagnetic resonance to be discussed below.

Although both phenomena are thus essentially identical, ferromagnetic resonance absorption is much stronger than the paramagnetic effect. Moreover, there is an important qualitative difference as regards the influence of the immediate environment of the atomic magnets. If this environment differs for individual magnets (e.g. in respect to the magnetic moment of neighbouring atomic nuclei) this will manifest itself, as we shall see, in a splitting or broadening of the paramagnetic resonance line. With ferromagnetic resonance, however, individual differences of environment are not noticeable since the atomic magnetic moments are all tightly coupled, and therefore these differences are averaged out.

Apparatus for measurements of paramagnetic resonance

A diagram of the set-up used for our investigations of paramagnetic resonance is given in *fig. 3*. As explained in the foregoing, paramagnetic resonance may be expected when atomic magnets in an external magnetic field are subjected to a transverse high-frequency field. This is done in the present apparatus by introducing the specimen into a resonant cavity T, which is subjected to a variable magnetic field B [9]). The resonant cavity is tuned to the frequency of a microwave generator (klystron K); the output from the klystron is transmitted to the resonant cavity via a waveguide G. The resonant cavity is tuned and matched to the waveguide in the absence of the external magnetic field. The klystron output is then used to keep the resonant cavity in oscillation, and only a weak signal is reflected into the waveguide. This being done, the magnetic field is increased until paramagnetic resonance sets in. This becomes manifest in a change in the tuning and matching of the resonant cavity. As a result the reflected signal increases. By means of a directional coupler R_2 the reflected signal is first separated from the much stronger wave travelling directly from the klystron. It is then detected, in our case by a sensitive superheterodyne method.

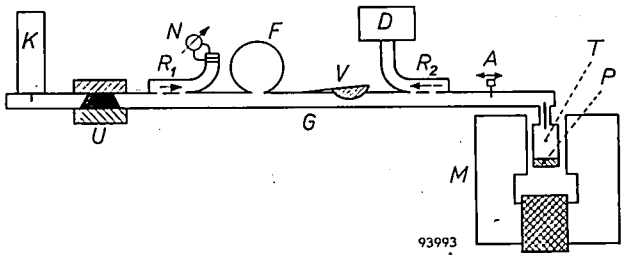


Fig. 3. Experimental set-up for observing paramagnetic resonance. K, klystron, which generates a wave of constant frequency f in the waveguide G. T, resonant cavity placed in the field of electromagnet M. P, specimen in resonant cavity. R_2, directional coupler, which conducts to detector D the signal reflected from T via matching stub A. The amplified detector output is automatically recorded as a function of the external field B. U, ferrite isolator, which prevents the reflected wave from affecting the klystron. The signal fed to T is controlled by a wavemeter F and a level-meter N with directional coupler R_1. V is a variable attenuator.

The reflection coefficient γ of a resonant cavity at the end of a waveguide depends on the damping and detuning and on the strength of the coupling between resonant cavity and waveguide, according to the formula:

$$\gamma = \frac{(1/Q_1) - (1/Q_2) - 2j(\Delta\omega/\omega_0)}{(1/Q_1) + (1/Q_2) - 2j(\Delta\omega/\omega_0)}. \quad \ldots\ldots \quad (9)$$

Here Q_1 is the "external" figure of merit of the resonant cavity (being the value of Q when only the losses through the coupling hole are taken into account); Q_2 is the figure of merit of the resonant cavity without coupling hole; $\omega_0 = 2\pi f_0$, where f_0 is the resonance frequency of the resonant cavity without external field; $\Delta\omega = 2\pi(f - f_0)$, in which f is the imposed frequency.

A paramagnetic specimen P placed in the resonant cavity and subjected to a suitable magnetic field will exhibit paramagnetic resonance (absorption and associated dispersion). This causes a change in the damping (Q_2) and in the tuning ω_0 of the resonant cavity, and hence also in $\Delta\omega$. In most cases only the paramagnetic resonance absorption is of interest, i.e. the increase in damping. According to formula (9), however, the reflected wave is in general dependent both on the change in the tuning (i.e. in $\Delta\omega$) and on the change in the damping (Q_2). The reflected waves, due to the increased damping or to detuning, differ in phase. By suitably choosing the operating point (Q_1; $\Delta\omega$) or by introducing an extra wave with the aid of the matching stub A (fig. 3), one can arrange for the detector to receive a wave which depends only on the damping.

[8]) H. G. Beljers and J. L. Snoek, Gyromagnetic phenomena occurring with ferrites, Philips tech. Rev. **11**, 313-322, 1949/50. H. G. Beljers, Amplitude modulation of centimetre waves by means of ferroxcube, Philips tech. Rev. **18**, 82-86, 1956/57 (No. 3). H. G. Beljers, The application of ferroxcube in unidirectional waveguides and its bearing on the principle of reciprocity, Philips tech. Rev. **18**, 158-166, 1956/57 (No. 6).

[9]) See also the article by H. G. Beljers and J. L. Snoek quoted under [8]).

The observations are carried out by leaving the frequency of the klystron constant and by examining the indication of the detector as a function of the magnetic field B, which is slowly varied about a specific value. This function, the "B spectrum", is made visible on an oscilloscope by modulating B at low frequency (equal to the frequency of the oscilloscope time base) and by making the vertical deflection on the oscilloscope proportional to the signal from the detector. *Fig. 4* shows a photograph of the complete apparatus.

In special cases much lower frequencies have been used, for example $f = 10^6$ c/s ($\lambda = 300$ m). This is possible if the spectra concerned are narrow (small Δf with constant B), that is to say if a single, narrow resonance line is to be examined, as in the case of organic radicals.

The normal method of calibrating the field is to record a comparison spectrum for a substance of known paramagnetic properties.

In view of the present wide application of paramagnetic resonance, Philips are now marketing a

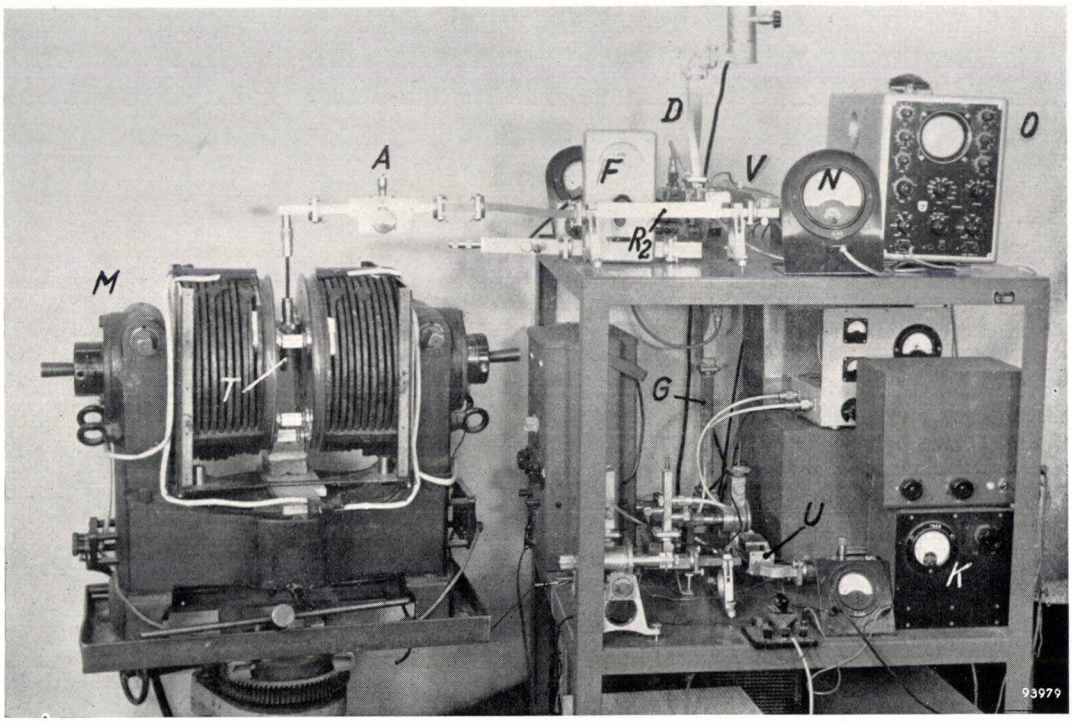


Fig. 4. Set-up used in the Philips laboratory at Eindhoven. *O* is an oscillograph. Other letters as in fig. 3.

For studying a given resonance spectrum the total width Δf of the spectrum (B remaining constant) or the corresponding width ΔB (f remaining constant) should be small with respect to the average value of f or B respectively. As already pointed out, this is why it is necessary (apart from the intensity of the phenomenon) to work with high frequencies and hence also with strong fields. In most cases the average value chosen for B is between 0.02 and 2 Wb/m^2 [200-20 000 gauss], the corresponding frequencies being 6×10^8 c/s and 6×10^{10} c/s ($\lambda = 50$ cm and 0.5 cm respectively). The preferred wavelengths are $\lambda = 1.25$ or 3 cm. Shorter waves and stronger fields are too difficult to produce experimentally to be suitable for day-to-day investigations.

range of components with which it is possible to build apparatus as in fig. 3 for frequencies up to 10 000 Mc/s.

Paramagnetic resonance spectra of Mn^{2+} in solids

As discussed in the introduction, paramagnetic resonance generally occurs not at a single frequency f_r, but in a certain frequency range in the neighbourhood of this frequency. A spectrum is found which has several components, each of which has a specific width and may also have an individual structure.

Let us now examine how this splitting process and broadening of the components come about. We have hitherto assumed that all atomic magnets (e.g. the resultant of the spins of the electrons of one ion) are subjected to the same field, namely to the

external field B. Frequently the atomic magnets are subjected not only to the field B but also to an extra field b, the origin of which is discussed below. According to (1) and (5) we know that resonance then occurs if [10])

$$f = g\frac{e}{4\pi m}(B+b), \quad . \quad . \quad . \quad (10a)$$

i.e.,

$$B = \frac{f}{ge/4\pi m} - b. \quad . \quad . \quad . \quad . \quad (10b)$$

The field b, and therefore the shift in the resonance spectrum, may not be the same for all the resonating atomic magnets, so that splitting or broadening of the lines may occur.

Assume for the sake of simplicity that the atomic magnet is a paramagnetic ion, whose paramagnetism arises from a single electron. The extra magnetic field b to which this electron is subjected may have several causes. We shall consider three of them:

a) the field due to a magnetic moment of the atomic nucleus;
b) fields due to neighbouring paramagnetic ions;
c) fields due to the movement of the electron through the electrostatic field of the crystal lattice (caused by nuclei and electrons).

[10]) The suffix to the symbol g will now be dropped since the electrons in a crystal lattice have a value of g (generally ≈ 2) which may differ from the value $g_e = 2.0023$ for a free electron.

In the following we shall deal first with a special case, that of divalent manganese ions (Mn^{2+}) dispersed randomly in a crystal lattice. We shall then go on to examine other applications of the investigation of paramagnetic resonance.

Effect of nuclear spin; hyperfine structure

The Mn^{2+} ion contains five electrons in an incompletely occupied electron shell, so that $S = \frac{5}{2}$ (in the ground state). If these electrons were subjected solely to the influence of the external field B, there would be six equidistant energy levels, the distance between which would be proportional to B. A representation of these energy values is given in *fig. 5a*. Resonance occurs according to (4) when $hf = \Delta E$, where f is the constant frequency of the klystron. Since the energy levels are equidistant, this condition is fulfilled for all transitions by one and the same value of B. Consequently the resonance would occur at only one value of B, and the spectrum would consist of a single absorption line, as represented below in fig. 5*a*.

In reality the situation is more complicated, since the nucleus of the manganese atom also has a magnetic moment. Manganese consists of only one stable isotope, the mass number of which is 55. The nucleus of this isotope has a spin given by its spin number $I = \frac{5}{2}$. The corresponding magnetic moment is $3.461(e/2m_p)(h/2\pi)$, where m_p represents the proton mass. The electrons that contribute to the para-

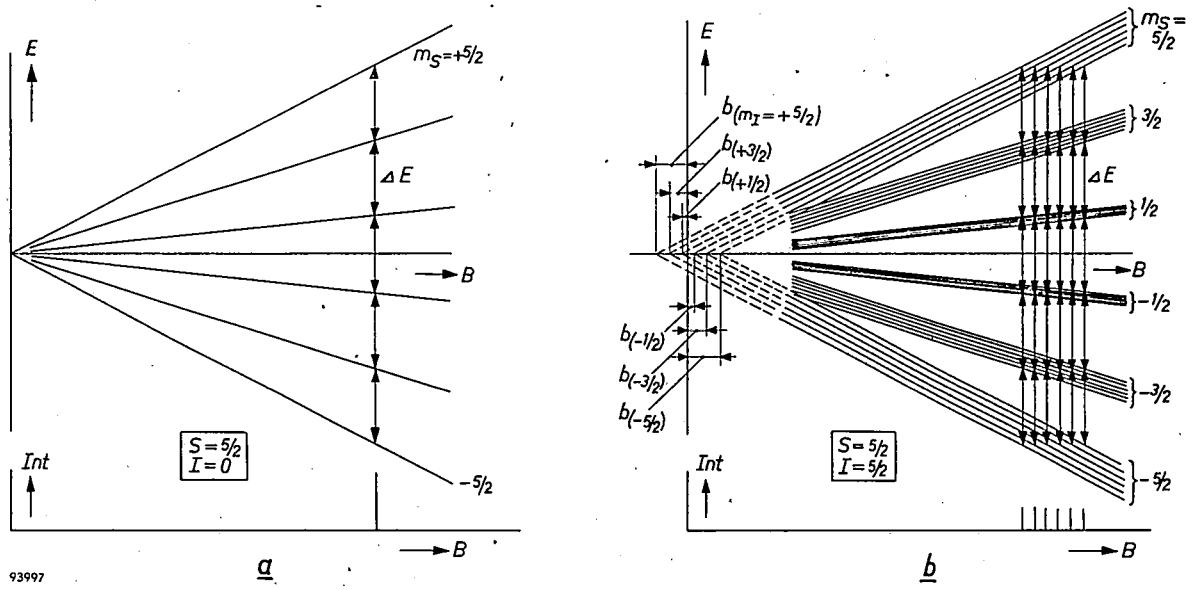


Fig. 5. *a*) Energy-level scheme of an ion with $S = \frac{5}{2}$ in a magnetic field B; below, the resulting resonance spectrum (intensity *Int* as a function of B), which in this case consists of a single line.
b) The same for an ion with $S = \frac{5}{2}$ which moreover has a nucleus of spin number $I = \frac{5}{2}$ (Mn^{2+}). Each of the levels of (*a*) is split into six sub-levels, corresponding to the six values of the field $B+b$. The 30 possible transitions give rise to six resonance peaks, as shown below.

magnetism are subjected both to the influence of the external magnetic field and to that of the nuclear field. Now, in a manner analogous to that described for electrons, a nucleus with spin number I has altogether $2I+1$ possible positions with respect to the field B. As a result, the parallel component b of the extra nuclear field, which acts on the electrons, also has $2I+1$ possible values. Each of the $2S+1$ levels of the paramagnetic ion is thus split into $2I+1$ levels (fig. 5*b*). Since $S=\frac{5}{2}$ and $I=\frac{5}{2}$, this gives $6\times 6=36$ levels. Since transitions occur only between the adjacent levels of fig. 5*a*, there are in all $6\times 5=30$ transitions between the split levels of fig. 5*b*, the energy differences relating to any one value of b being equal. In consequence of this the resonance condition is now fulfilled for six values of B, so that the resonance line in fig. 5*a* is in reality split into six components, as illustrated below in fig. 5*b*.

This, of course, provides information only on those electrons belonging to the outer, non-closed electron shells, for the electrons in closed shells do not contribute to the paramagnetism. Now it is precisely for the outer-shell electrons that the extent of the orbits, and hence the average field to which they are subjected by their "own" nucleus, is affected by the environment of the ion. Paramagnetic resonance can therefore also provide information on this environment. *Fig. 6* shows the paramagnetic resonance spectra of Mn^{2+} ions incorporated in different crystal lattices [12]). All these crystals contain divalent ions. A small fraction of these (0.1—0.01%) are replaced by Mn^{2+} ions. It can be seen from fig. 6 that the splitting of the resonance

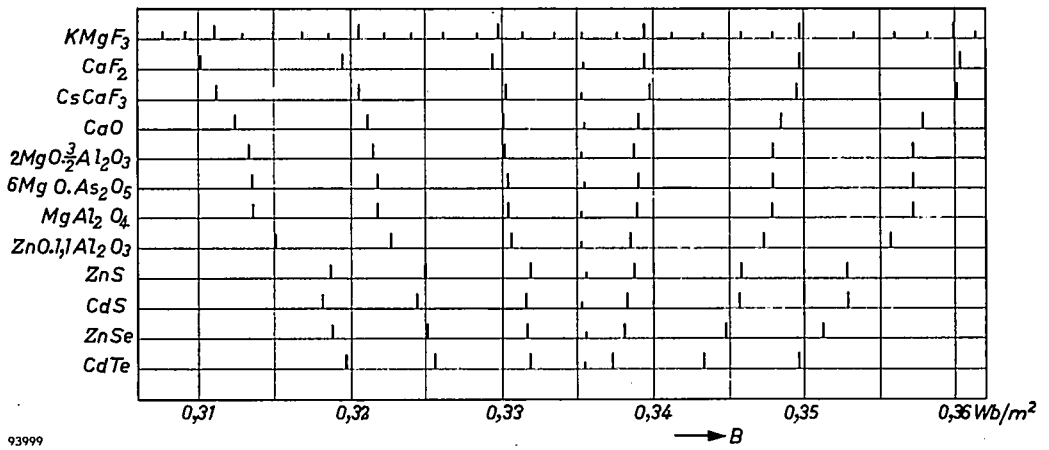


Fig. 6. The resonance spectrum (only the *position* of the peaks is shown) of Mn^{2+} in various crystal lattices. The magnitude of the splitting depends mainly on the nature of the negative ion.

This splitting into components (hyperfine structure [11])) occurs in many other ions besides the example given of Mn^{2+}. (About half the stable isotopes of all elements possess a nuclear spin.) The magnitude of the splitting is determined by the average field excited by the atomic nucleus at the position of the electrons involved in the resonance. This average field depends on the nature of the nucleus and on the extent of the electron orbit: the further away the electron is from the nucleus, the smaller is the splitting. A study of this splitting process therefore yields information on the spatial distribution of the electrons.

line of Mn^{2+} does, in fact, differ considerably in different environments. From the results of measurements on three different fluorides ($KMgF_3$, CaF_2 and $CsCaF_3$), and on two sulphides (CdS and ZnS), of widely differing lattice constants, it appears that the latter play hardly any role in this process. The magnitude of the splitting is almost entirely determined by the nature of the negative ions surrounding the Mn^{2+} ions. With F, O, S, Se and Te the average magnitude of splitting in the B spectrum is respectively 98, 80-90, 69, 65 and 59×10^{-4} Wb/m^2.

Now it is known that in the above series of compounds (fluorides, oxides etc.) the bond is the most ionic in nature in the fluorides, and that in the order given it becomes gradually more homopolar

[11]) In the theory of atomic spectra the term fine structure is used when referring to a multiple spectral line (e.g. the D_1D_2 doublet of sodium) produced by the splitting of an energy level as a result of *spin-orbit* interaction; the term hyperfine structure relates to a multiplet produced by interaction with the *nuclear spin*. The hyperfine structure is generally small as compared with the normal fine structure. In the case of paramagnetic resonance this terminology is retained although, here, the effect of the nuclear spin may be large compared with other effects.

[12]) J. S. van Wieringen, Paramagnetic resonance of divalent manganese incorporated in various lattices, Disc. Faraday Soc. **19**, 118-126, 1955.

in nature. The case of the more homopolar compounds can be characterized by saying that the outermost electrons of the cation (Mn^{2+}) are partly associated with the electron clouds in the surrounding anions, and thus, on an average, are farther away from the Mn nucleus. The differences in the magnitude of the splitting in fig. 6 are accordingly in good agreement with what was known qualitatively of the nature of the bond between the ions of the series listed above, and moreover have provided useful quantitative information.

It can be seen from this that the study of the structure of paramagnetic resonance lines is a valuable tool in investigations into the nature of the bonding in crystals. It is used in an analogous way to analyse the bonds in complex molecules: the unsaturated bond present in a free radical, which functions as an atomic magnet, has been localized near certain atoms in the radical in many cases by studying the paramagnetic resonance spectrum.

Effect of the magnetic fields of neighbouring paramagnetic ions; line broadening

Apart from the external magnetic field, and possibly the field of the nuclear spins, another field may be operative: this is the field originating from neighbouring paramagnetic ions, particularly from those of the same kind as the resonating ion. This field also contributes to the term b in the resonance condition (10). Since, however, it originates from a large number of similar ions, which all occupy different attitudes in space and are at different distances from the ion concerned, the result in this case is not a splitting into a relatively small number of lines (as in the case of the nuclear spin), but a splitting into so many components that only a broadening of the resonance line is observed. The greater the concentration of the paramagnetic ions, the smaller is their average distance from each other and the stronger are the magnetic fields they exercise on their neighbours, and hence the greater is the broadening of the resonance line. This can be seen from *fig. 7*, which shows the paramagnetic resonance spectrum of Mn^{2+} added to ZnS, for varying percentages of manganese [13]). With increasing manganese content the spectrum changes from six more or less separate lines into a single broad maximum, exhibiting very little structure. This is also the reason why it is, in general, not possible to separate the resonance spectrum into components in a pure compound of paramagnetic ions (e.g. MnS). The components only become visible when the paramagnetic ions have been strongly diluted by the admixture of non-magnetic ions, such as Zn^{2+} (see fig. 7).

Effect of electron motion in the electrostatic crystalline field

The influence of the crystalline field becomes manifest in a splitting of the energy levels, which appears even in the absence of an external magnetic field. A proper analysis of this phenomenon involves rather complicated and detailed considerations, which cannot be pursued here. It will be enough for our purposes to discuss merely the broad outlines.

It was assumed in the foregoing that the magnetic moment of atomic magnets (i.e. in our case the Mn^{2+} ions) is due solely to electron spins. This assumption is not of general validity, since the orbital motion of the electrons also contributes to the magnetic moment. In the case of free paramagnetic ions and atoms the magnetic moment is the resultant of a spin and an orbital moment which are of the same order of magnitude.

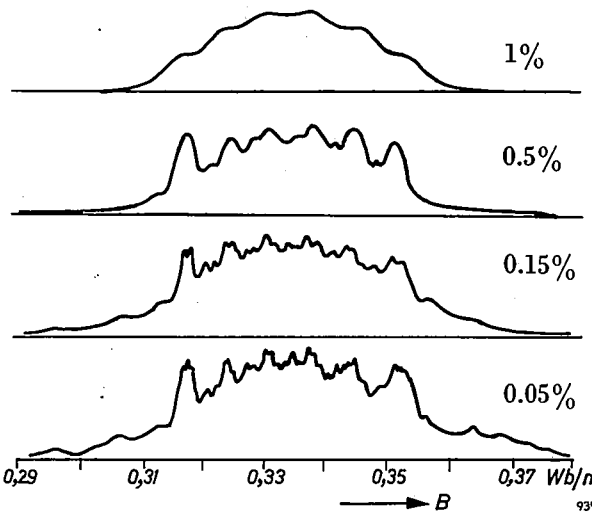


Fig. 7. Effect of the paramagnetic ion concentration on the spectrum. The spectrum of hexagonal ZnS (wurtzite) with the addition of various amounts (atom %) of Mn.

A paramagnetic ion or atom in a crystal lattice, however, is subjected to strong electrostatic fields. This has two important effects on its magnetic properties. In the first place the orbital moment is almost entirely destroyed, because the electrons move in their orbits in different directions (a quantum-mechanical effect, described in small print below) and this makes the resultant orbital current vanishingly small. The magnetic moment of the ion (and the same applies to electrons in paramagnetic centres) is therefore identical with the spin moment. This fact has already been employed in the deriva-

[13]) J. S. van Wieringen, Influence du traitement mécanique sur la résonance paramagnétique du manganèse dans les poudres de sulfure de zinc, Physica **19**, 397-400, 1953.

tion of formula (5). If, however, we wish to consider the situation more precisely, we must take into account the presence of a certain field due to the orbital moments, which causes the total magnetic moment of the incorporated ion to differ somewhat from the spin moment. Consequently the factor g_e (= 2.0023) in (5), which applies to a free electron, must be replaced by another factor g. This causes a shift in the field B at which resonance occurs; see equations (1) and (10), in which this is taken into account.

The second effect of incorporating the ion in a crystal lattice is that the orbital moment, in so far as it is still present, becomes anchored to the electrostatic field of the crystal. It can therefore no longer move freely in space, unlike the spin moment, which is almost entirely free. The spin moment is, however, only weakly coupled to the orbital moment by the spin-orbit interaction; the coupling with the crystal lattice is therefore small.

Owing to the orbital motion being bound to the crystalline field, the contribution of the residual orbital moment to the total magnetic moment of the atomic magnet depends as a rule on the orientation of the crystal in the external field B. The spectrum is therefore also dependent on this orientation; in other words, the paramagnetic resonance spectrum is not, generally speaking, isotropic with respect to the external field. The magnitude of the difference between g and g_e, and the symmetry of the anisotropy depend on the strength and the symmetry of the crystalline field and on the structure of the ion or atom (or other centre) functioning as atomic magnet.

We now come to the earlier mentioned splitting of the paramagnetic resonance lines which may be caused by the crystalline field. Spin-orbit interaction is a consequence of the motion of the electron in the electric field of the crystal lattice giving rise to a local magnetic field at the position of the electron (this field attempts to orient the magnetic moments of spin and orbit in opposite directions). The other electrons and nuclei in the lattice are in a state of motion with respect to the electron concerned, and therefore function as electrical currents which give rise to a magnetic field. This extra magnetic field can be taken into account, as in (10), by adding a field b to the external field B. The strength of the spin-orbit interaction, hence the magnitude of b, depends on the angle between the spin moment and the orbital moment. It may therefore vary for different spin orientations, and moreover it again depends on the attitude of B with respect to the crystalline field. Spin-orbit interaction, then, can give rise to line splitting, the magnitude of which depends on the orientation of the crystal in the external magnetic field (and again, of course, on the structure of the ion, atom or other centre acting as atomic magnet); see *fig. 8*.

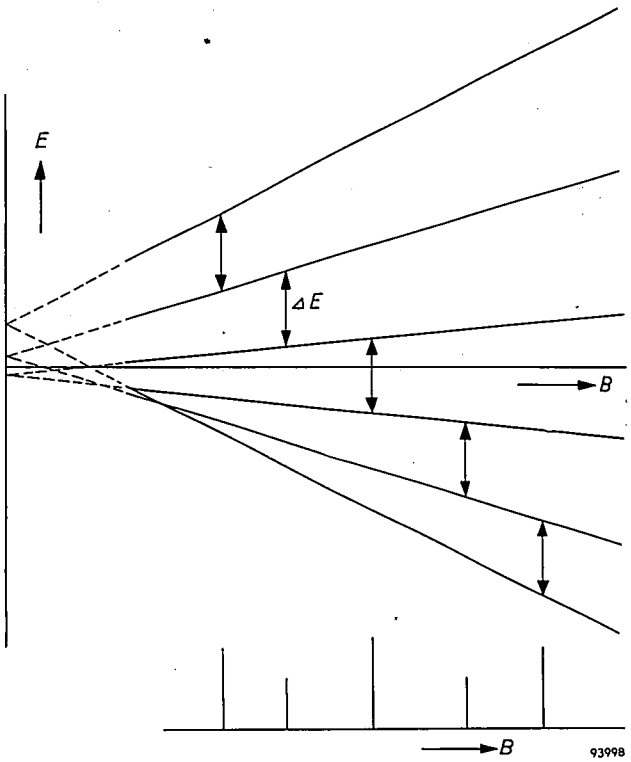


Fig. 8. Illustrating the effect of electron motion in the electrostatic field of the crystal. In a non-cubic crystal different orientations of the spin with respect to the crystallographic axis correspond to different energies. In the case of $S = \frac{5}{2}$ there are three energy levels, corresponding to $m_s = \pm\frac{1}{2}, \pm\frac{3}{2}, \pm\frac{5}{2}$. In a magnetic field each of these levels is split into two. Below, the resultant resonance spectrum (five lines). The effect of nuclear spin is neglected here.

This line splitting is only possible when the spin number of the atomic magnet is greater than $\frac{1}{2}$. Where the spin number is $\frac{1}{2}$ there are only two levels and only one transition can take place between them, and therefore no splitting pattern can appear. Larger spin values can give rise to a complicated spectrum.

The magnitude of the splitting depends not only on the attitude of the crystal in the magnetic field but again on the nature of the crystalline field in which the ion is situated. The lower the symmetry of the crystalline field the greater the splitting. For cubic crystals the splitting is very often zero, or at least very small.

We shall illustrate this with an example [13]), again taking the Mn^{2+} ion. *Fig. 9a* shows the resonance spectrum of wurtzite (hexagonal ZnS) to which a small percentage of Mn (0.05 percent) has been added. The spectrum is a very complicated one. The cubic modification of ZnS (sphalerite) shows the

simple spectrum with six hyperfine-structure components, fig. 9*e*. Now the paramagnetic resonance spectrum of Mn in hexagonal ZnS is found to change considerably when the specimen is deformed. This can be seen from fig. 9*b*, *c*, *d*, which show the spectrum after compression of the specimen at different

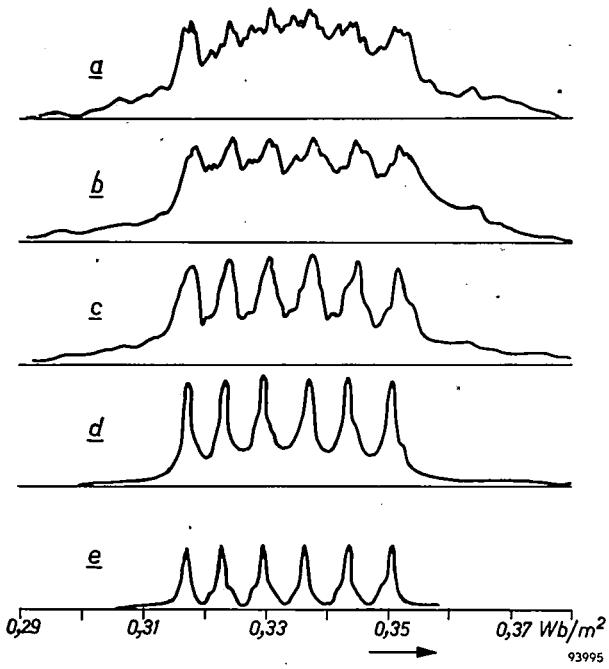


Fig. 9. Example of how the symmetry of the crystal lattice affects the resonance spectrum. In *a*) and *e*) are shown the resonance spectra of Mn^{2+} (0.05 %) in wurtzite (hexagonal ZnS) and in sphalerite (cubic ZnS), respectively. The other spectra show how, under an applied pressure, the crystal structure changes from hexagonal to cubic. The relevant pressures were: *b*) 750 kg/cm², *c*) 1500 kg/cm², *d*) 3500 kg/cm².

pressures. The spectrum gradually changes during this process to that of cubic ZnS. The explanation of this phenomenon is that hexagonal ZnS is metastable at room temperature. The cubic structure represents the stable state. A relatively small pressure is sufficient to convert the hexagonal form into the stable cubic form; the conversion is more complete the higher is the pressure, more crystals then changing from the hexagonal to the cubic form. This process can be followed closely by studying the paramagnetic resonance of incorporated Mn.

The transition from the hexagonal to the cubic lattice is demonstrated by a crystal model in *fig. 10*.

It is instructive to examine how the quantum-mechanical description accounts for the effect of the crystalline field on the orbital moment. We shall start from the quantum-mechanical explanation of the bonding in the hydrogen molecule ion H_2^+.

With this ion the electron is associated now with the one and now with the other hydrogen nucleus. If we represent this by two wave functions ψ_1 and ψ_2, we can combine these into a symmetrical function $\psi_1 + \psi_2 = \psi_s$ and an anti-symmetrical function $\psi_1 - \psi_2 = \psi_a$. The chance of an electron being found at a particular point in space is given by $|\psi|^2$ at that point. In the case of ψ_s the most probable position of the electron is between the two nuclei, for ψ_a the most probable position is outside them. If the nuclei are widely spaced, the energy of both wave functions is equal ($E_s = E_a$). If the distance between the nuclei is decreased, then $E_a > E_s$ and the symmetrical wave function represents the energetically more favourable state. The effect of introducing a perturbation in the form of an external electrical field which exerts a force on the electron, e.g. parallel to the line connecting the nuclei, and directed from nucleus 2 to nucleus 1, is to change the wave functions in this field. The energetically more favourable wave function can now be represented approximately by a linear combination of the original ψ_s and ψ_a:

$$\psi = A\psi_s + B\psi_a,$$

where $A \gg B$. It is said that the function ψ_s is mixed with a little of ψ_a. The result of this is that ψ_1 and ψ_2 in the expression for $|\psi|^2$ no longer occur symmetrically, which can be interpreted by saying that there is now a greater probability of finding the electron near nucleus 1 than near nucleus 2. The external field "blows", as it were, the electron towards nucleus 1.

Something of the kind occurs when a paramagnetic ion is placed in a crystalline field. If by ψ_1 and ψ_2 we now understand wave functions corresponding respectively to an anti-clockwise and clockwise motion of the electron, we again have the combinations ψ_s ($= \psi_1 + \psi_2$) and ψ_a ($= \psi_1 - \psi_2$). One of these is energetically more favourable, because the spatial distribution of the electron it represents is more favourable with respect to the crystalline field. The orbital moments corresponding to both ψ_s and ψ_a are zero (i.e. anti-clockwise and clockwise motion are equally probable). The spin-orbit interaction remains, however. It has the character of a perturbation, which gives rise to another energetically most favourable state, for which the wave function is again approximately given by $A\psi_s + B\psi_a$, where, e.g., $A \gg B$; in other words, there is again a mixing of wave functions. The result is that ψ_2, for example, prevails over ψ_1. The probability of clockwise motion of the electron is then greater than that of anti-clockwise motion, and therefore a residual orbital moment exists.

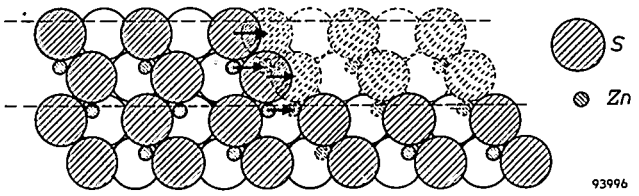


Fig. 10. Cross-section of a crystal of hexagonal ZnS parallel to a plane containing the hexagonal axis and a two-fold axis (left half of diagram) and cross-section of face-centred cubic ZnS crystal parallel to a (110) plane (right). Both lattices are close-packed structures of sulphur ions. The tetrahedral sites are occupied by zinc ions. The sulphur and zinc ions lie in layers perpendicular to the plane of the drawing. A comparison of the left and right sides shows that the bottom and next-to-bottom layers in the figure are identical. The difference between the hexagonal and the cubic structure appears in the third layer from the bottom.

Under the influence of an applied pressure (at room temperature) hexagonal ZnS changes to cubic ZnS. This occurs by a glide process, in which each third layer glides over the one immediately below it. The drawing shows two glide planes (dashed lines). The arrows represent ion displacements during the glide process.

Since in this case the energy difference between the states ψ_s and ψ_a is of the order of 0.1 eV, and the spin-orbit interaction of the order of 0.001 eV, the perturbed state lies very close to the state of zero orbital moment. The resultant orbital moment is therefore always very small compared with that of a free atom or ion.

A particle that simultaneously rotates anti-clockwise and clockwise (with almost equal probabilities of either) corresponding, for example, to the function ψ_s, is of course unthinkable in classical mechanics. This is a typical illustration of the fact that the behaviour of electrons in atoms and ions is entirely different from that of macroscopic particles, and can be described only in terms of quantum mechanics.

Let us now briefly summarize the information that can be deduced from the investigation of paramagnetic resonance spectra.

The *magnitude* of the resonance absorption generally indicates the number of atomic magnets in the substance investigated.

From the *displacement* of the resonance spectrum (difference in g value) the orbital contribution to the magnetic moment can be deduced. With single crystals the anisotropy (with respect to the field) of the displacement provides information on the symmetry properties of the crystal lattice in the immediate neighbourhood of the atomic magnet.

The *line width* — if caused by atoms or ions of the same kind — yields information on the mutual spacing of the atomic magnets.

Line-splitting can have several causes. So-called hyperfine structure is an effect of nuclear spin, while so-called fine structure is due to the crystalline field. We can distinguish between these two possibilities by observing the ratio of intensities of the line components. In the case of hyperfine structure, the number of components provides information on the nucleus of the atomic magnet (or on the nuclei in the immediate vicinity). The magnitude of the splitting allows deductions to be made regarding the spatial distribution of the electrons of the atomic magnet concerned, and possibly regarding the inter-atomic bonds. If, on the other hand, the splitting is caused by the crystalline field, it tells us something about the spin of the atomic magnet and (in single crystals) about the symmetry of the local crystalline field.

Other investigations of paramagnetic resonance

Apart from the investigations described above, which relate to Mn^{2+} ions, we shall now describe a few other interesting cases, some of which have been examined in this laboratory.

F centres of alkali-halides

Alkali-halides (NaCl, KCl, etc.), when subjected to X-radiation, show an absorption band in or near the visible part of the spectrum. KCl, for example, becomes violet because an absorption band is produced in the green. It has been found from various experiments that this absorption band is attributable to a halogen ion being replaced by an electron on some lattice sites [14]). These sites are called F centres. An old dispute in this connection concerned the question of how strongly the electron is localized. Does it remain inside the vacancy left by the halogen ion or is it distributed over the six surrounding positive alkali ions?

This dispute was settled by experiments carried out by Kip, Kittel, Levy and Portis [15]). Since the F centre contains a single electron, it is paramagnetic. It exhibits paramagnetic resonance, having a spectrum, as expected, consisting of one line. One of the observables is the displacement of the line, from which it follows that, for KCl, $g = 1.995$ (against $g_e = 2.0023$ for free electrons). This points to a relatively large moment, which can be better explained by the dispersed model than by the localized one. An additional confirmation of the dispersed model follows from the line width, which is too large to be explained in terms of the magnetic interaction of centres dispersed at random over the lattice. At first it was deduced from this that the F centres occurred in clusters. It later appeared, however, that the line width was caused by the electron interacting not only with the six nearest-neighbour alkali nuclei but also with the twelve next-nearest-neighbour halogen nuclei. Since there are so many nuclei involved, the interaction in this case does not lead to a clearly visible hyperfine structure, but to a line broadening with unresolved separate components. The fact that the broadening in this case is caused by interaction with nuclear spins was convincingly demonstrated by making use of other alkali isotopes. These have a different nuclear spin and a different interaction. It was possible to give a quantitative explanation of the different line width then observed.

Centres in irradiated quartz crystals

Griffiths, Owen and Ward [16]) found paramagnetic resonance in quartz crystals subjected to X-radiation. The resonance line showed a hyperfine struc-

[14]) J. H. de Boer, Electron emission and adsorption phenomena, Univ. Press, Cambridge 1935. See also E. J. W. Verwey, Electronic conductivity of non-metallic materials, Philips tech. Rev. **9**, 46-53, 1947/48.

[15]) C. A. Hutchison, Phys. Rev. **75**, 1769, 1949.
A. F. Kip, C. Kittel, R. A. Levy and A. M. Portis, Phys. Rev. **91**, 1066, 1953.
N.W. Lord, Phys. Rev. **105**, 756, 1957 (No. 2).

[16]) J. H. E. Griffiths, J. Owen and I. M. Ward, Nature **173**, 439, 1954.
Mary C. M. O'Brien, Proc. Roy. Soc. London A **231**, 404, 1955.

ture of six lines, and the resonance field B was found to depend on the attitude of the crystal in the external magnetic field (anisotropy of g). The hyperfine structure appeared to be due to aluminium. The investigation led to the conclusion that the centres formed were localized at Al ions which occur as impurities in SiO_2. The examination of the anisotropy showed that the charge cloud of the centre is oriented along the lines connecting oxygen atoms. A detailed description has been given by Mary O'Brien.

Donors and acceptors in semiconductors

Some crystalline substances, such as Si and SiC which are almost perfect insulators in the pure state, can be made into semiconductors by the addition of a very small percentage of donors or acceptors, i.e. foreign atoms which either give up an electron to the crystal lattice or absorb an electron from the lattice (in the latter case conductivity is obtained by displacement of the electron hole in the lattice). At very low temperatures, the electron or the electron hole does not leave the donor or the acceptor, as the case may be, but remains localized in its vicinity. For this experiment Si must be placed in liquid helium (—269 °C) [17]) and SiC in liquid nitrogen (—190 °C) [18]). In that state the donor or the acceptor atoms are generally paramagnetic. This appears from the occurrence of a paramagnetic resonance line, in which, moreover, a hyperfine structure is found. At higher temperatures this line vanishes, because the electrons or holes are then released from the donors or acceptors and are able to move more or less freely through the lattice. Resonance without hyperfine structure may then occur, as in fact has been observed in the case of Si.

Summary. After a brief description of the phenomenon of paramagnetic resonance, the quantum-mechanical theory of the phenomenon is discussed. Reference is made to the relation between paramagnetic resonance and the Zeeman effect, and to the relationship with ferromagnetic resonance in ferromagnetics. For observing paramagnetic resonance use is made of a resonant cavity, containing the substance under investigation and attached to a waveguide through which radiation of constant frequency is passed (e.g. $\lambda = 3$ cm), the cavity being placed in a constant magnetic field. The strength of this field is varied and the signal reflected from the resonant cavity is detected and recorded as a function of the field strength. This reveals a spectrum which usually shows several peaks, each of which has an individual structure. An extensive investigation has been made of the spectrum in the case of divalent manganese, Mn^{2+}, in various crystal lattices. Six components are usually found (hyperfine structure due to nuclear spin). The influence of the bonding in the crystal lattice (competition between ionic and homopolar bonding), of the manganese content (neighbouring ions) and of the symmetry of the crystal lattice (hexagonal and cubic ZnS) are discussed in the light of the experimental results. Mention is then made of some other results of paramagnetic-resonance investigations, namely the resonance of F centres of alkali-halides, of centres in irradiated quartz crystals and of donors and acceptors in semiconductors. The examples chosen give some idea of the information that can be derived from paramagnetic resonance spectra.

[17]) R. C. Fletcher, W. A. Yager, G. L. Pearson and F. R. Merritt, Phys. Rev. **95**, 844, 1954.
A. M. Portis, A. F. Kip, C. Kittel and W. H. Brattain, Phys. Rev. **90**, 988, 1953.

[18]) J. S. van Wieringen, Internationales Kolloquium über Halbleiter und Phosphore, Garmisch-Partenkirchen, 1956; to be published by F. Vieweg, Brunswick 1958.

AN X-RAY DIFFRACTION TUBE WITH ROTATING ANODE FOR 10 kW CONTINUOUS LOADING

621.386.26:548.73

X-ray tubes for use in diffraction analysis are usually of the sealed-off type, the anode being fixed and water-cooled. If such a tube has a copper anode and a true focus measuring 10 mm², say, it will be capable of taking a maximum power load of round about 1 kW. If other materials are used for the anode and the design otherwise remains unchanged, the maximum power will not be more than a few hundred watts.

For those kinds of diffraction analysis that may call for extremely long exposure times, it is of advantage to have tubes of much higher ratings available. An example of such work that comes immediately to mind is the study of certain phenomena by means of the *diffuse* scattering. One important type of investigation coming under this head is that into the kind of movement performed by certain atomic groups in some crystals when the temperature rises above a certain value (for example, the NO_3 radical in $NaNO_3$ starts rotating as the compound passes into the modification stable at high temperatures). The structure of mixed crystals also gives rise to diffuse scattering which is sometimes measurable (e.g. in NaCl-KCl). Another field in which the investigator is concerned with low intensities is the analysis of albumin structures.

Higher power ratings can be obtained by arranging for the water-cooled anode to rotate. Only a small number of X-ray diffraction tubes with a rotating water-cooled anode have been constructed [1]). A tube of this kind has been built in the Philips laboratories at Eindhoven in order to gain experience in the various design problems involved. The tube has a copper anode — copper radiation is the most commonly used in diffraction analysis — and can take a continuous load of 10 kW (40 kV at 250 mA). It is very compact in construction.

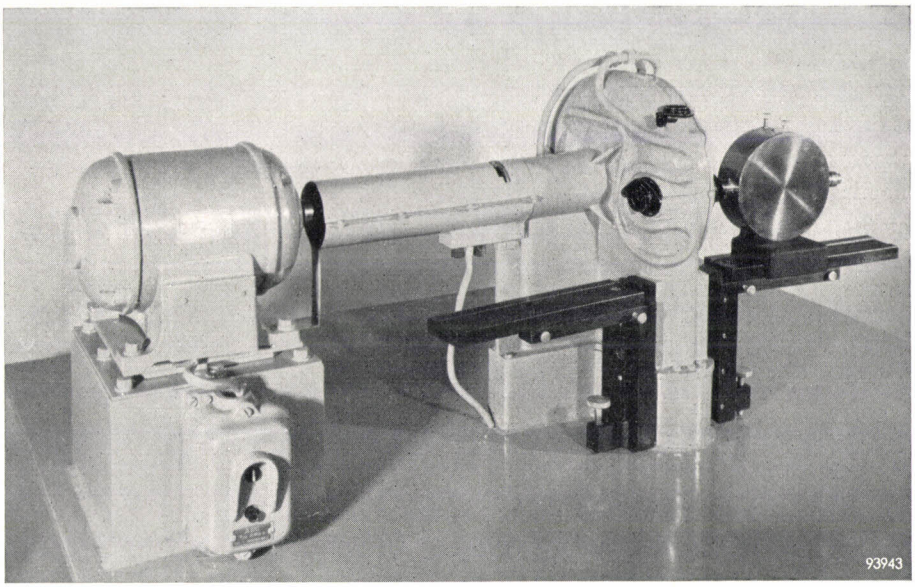


Fig. 1. A rotating-anode X-ray diffraction tube built in the Philips laboratories at Eindhoven; the rotating water-cooled anode can take a continuous load of 10 kW. The motor is on the left, the housing accommodating the anode disc is on the right, and on the extreme right, mounted on a slide, is an X-ray diffraction camera.

Fig. 1 is a view of the exterior of the tube. On the left is a three-phase motor, which drives the anode (at 920 r.p.m.) via a shaft with a universal joint; the housing in which the anode disc is accommodated is on the right, and on the extreme right is a camera mounted on a slide. There is also a camera slide on the motor side of the housing.

The focus is formed on the outer edge of the anode disc and has dimensions 1 mm × 10 mm (*fig. 2*). Three beryllium windows, 1 mm thick, allow the passage of radiation to the exterior; they are mount-

[1]) A. Müller, Brit. J. Radiol. **4**, 127, 1931.
R. E. Clay, Proc. Phys. Soc. **46**, 703, 1934.
A. Taylor, J. sci. Instr. **26**, 225, 1949.
W. T. Astbury, Brit. J. Radiol. **22**, 360, 1949.
E. A. Owen, J. sci. Instr. **30**, 393, 1953.

ed in black frames that may be distinguished in fig. 1. Through two of the windows the radiation leaves the housing in a horizontal direction; from the third it is beamed vertically upwards. Since the

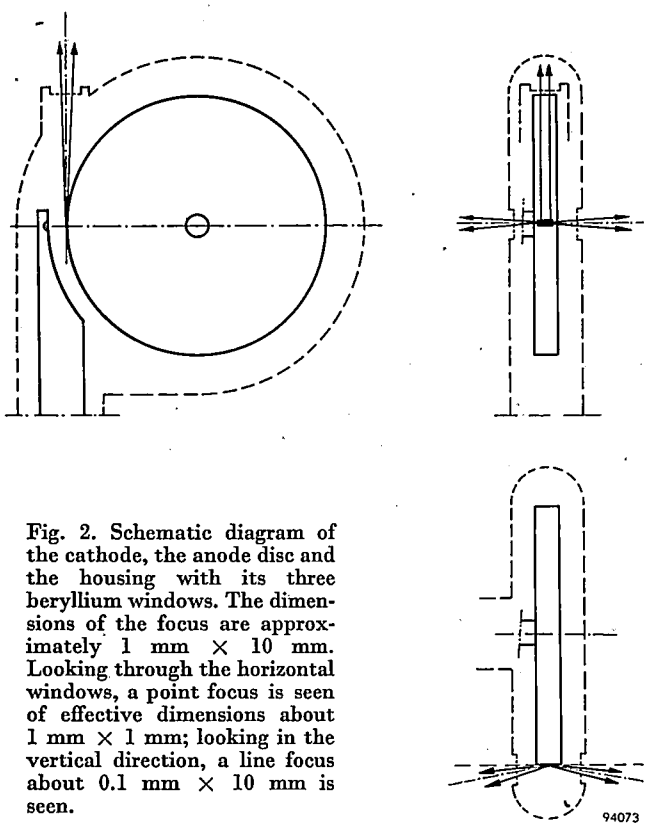


Fig. 2. Schematic diagram of the cathode, the anode disc and the housing with its three beryllium windows. The dimensions of the focus are approximately 1 mm × 10 mm. Looking through the horizontal windows, a point focus is seen of effective dimensions about 1 mm × 1 mm; looking in the vertical direction, a line focus about 0.1 mm × 10 mm is seen.

slides are almost parallel with the axis of the disc (at an angle of 6° to this axis), a point focus of approximately 1 mm × 1 mm is presented to the camera; looking through the upper window, one sees a line focus [2]) of approximately 0.1 mm × 10 mm. The windows can be covered on the outside by small lead shutters.

Fig. 3 shows a vertical section through that part of the tube containing the shaft and the anode disc. The plane of the drawing passes through the axis of the shaft (*A-A*). The housing is made up of a fixed portion *C* and a removable circular cover *E*. Removal of the cover allows the anode disc *D* and the shaft to be extracted from the housing for cleaning and maintenance purposes.

The shaft turns in ball-bearings (B_1 and B_2) and passes through five tight-fitting rings (numbered *1-5*) made of a synthetic rubber that is proof against oil and heat. These rings have an internal metal reinforcement, and they fit closely round the shaft and round the inside of the housing, so that they divide the enclosed space into a series of chambers.

Throughout that part of its length marked *a-a*, the shaft consists of two co-axial cylinders. The cooling water entering at inlet *F* passes into the chamber formed by rings *4* and *5*, and penetrates thence through holes *f* into *L*, the inner cylinder. From there the water flows through three radial pipes inside the disc into a runnel *J* which underlies the anode strip *K* throughout the periphery of the disc. The return path is via three other radial tubes into *H*, an annular space between the two cylinders just referred to; the water finally leaves the tube by way of holes *g*, the chamber formed by rings *3* and *4*, and outlet *G*. A high vacuum exists in the large compartment to the right of ring *1*, which makes a vacuum-tight fit around the anode shaft. The chamber between rings *1* and *2*, which contains ball-bearings B_1, leads to a vacuum reservoir connected to the backing pump. That between rings *2* and *3* separates the vacuum space from the cooling-water chamber, and serves to drain off any water that may leak through ring *3*.

Since the part of the housing in the vicinity of the focus is subjected to secondary-electron bombardment while the tube is in operation, it must also be cooled: this is necessary above all to protect the various rubber rings that serve as vacuum seals for the windows etc. The power to be dissipated from the housing when the tube is under full load is about

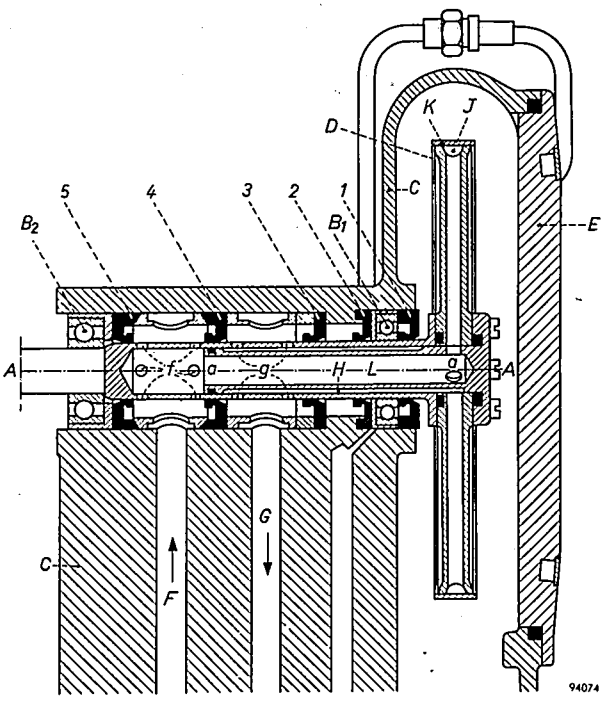


Fig. 3. Vertical section through the housing, the shaft and the anode disc. The drawing shows the bearings, seals and water-cooling arrangements for the rotating anode (one or two non-essential features have been omitted).

[2]) The use of the point focus and of the line focus is discussed in J. E. de Graaf and W. J. Oosterkamp, Philips tech. Rev. 3, 259, 1938.

1 kW. The cooling-water pipes fixed to the exterior of the housing may be seen in fig. 1.

A centrifugal switch mounted on the shaft cuts off the anode tension whenever the disc stops turning, whatever the reason. The switch is mounted inside the cylindrical housing, between the motor and the X-ray tube housing (see fig. 1).

The cathode assembly has a tungsten filament 0.3 mm thick, fixed to the lead-in wires by means of two small screws; it can therefore be changed very quickly. The life of the filament is about 200 hours. The cathode assembly is inserted into the tube from below, a vacuum-tight closure being made by means of a flanged screw-cap and a rubber ring. Part of the porcelain insulator carrying the supply leads to the cathode can be seen in the general view of *fig. 4* (owing to the mechanical design of the tube the anode has to be at earth potential, and hence the cathode assembly is at high tension). Beneath the insulator is a black metal sphere that serves to

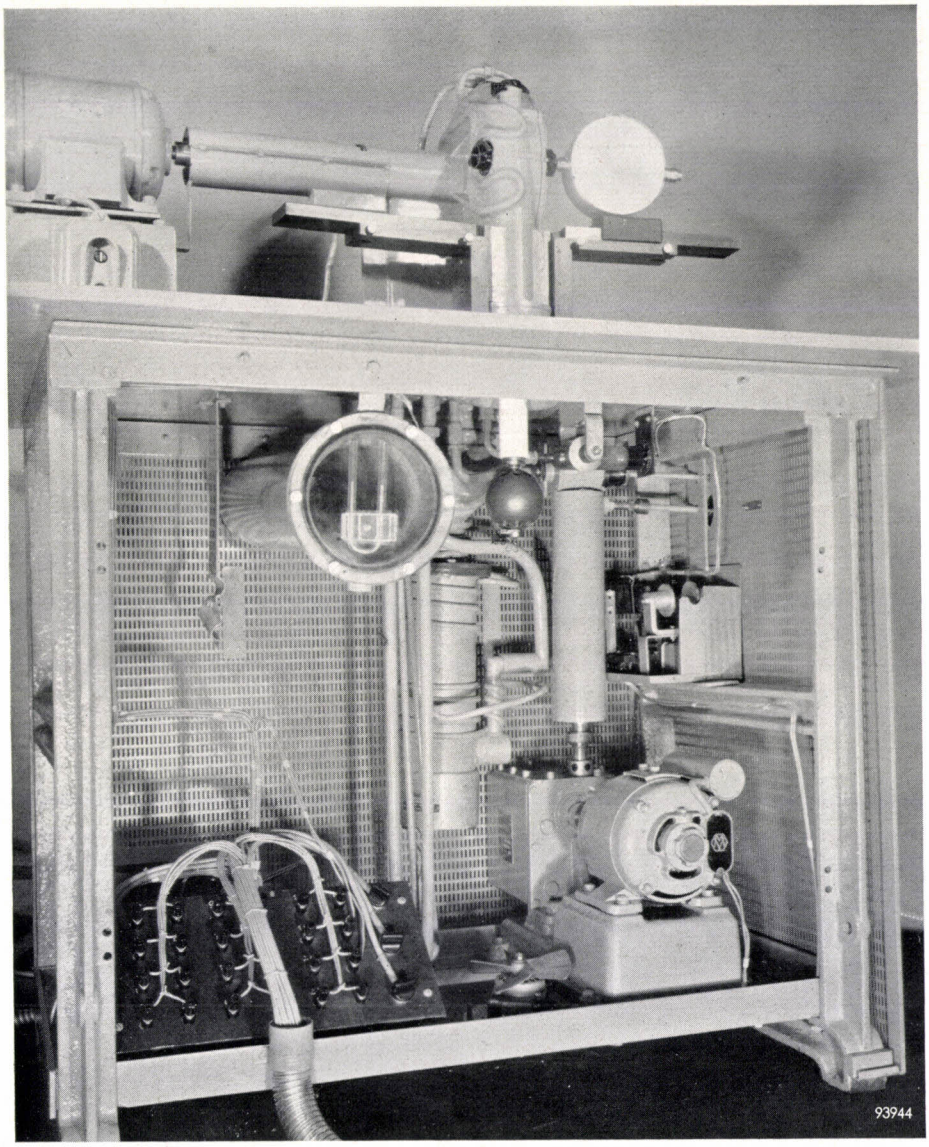


Fig. 4. The table on which the X-ray tube is mounted, and housing, underneath, the vacuum system and other auxiliaries. The rotary oil pump may be seen in the bottom right-hand corner (backing pressure 10^{-3} mm Hg). At the rear, in the middle, is the two-stage oil diffusion pump. It is possible with these pumps to obtain a vacuum of better than 10^{-5} mm Hg. Near the top of the compartment, centre, may be seen the lower end of the porcelain insulator which forms part of the cathode assembly. The metal sphere attached to the bottom of the insulator helps to dissipate heat; it is also the terminal for the high-tension cable (absent in this photograph). To the right of the insulator are the tube and magnet of a Penning manometer. The electrical part of this instrument is mounted half-way up the wall on the right. The large cylindrical container with end-window is the vacuum reservoir; inside it is a mercury manometer. The cable in the foreground goes to a separate control unit.

dissipate heat; the high-tension cable is attached to this sphere. Fig. 4 also shows the vacuum pumps and various other auxiliaries.

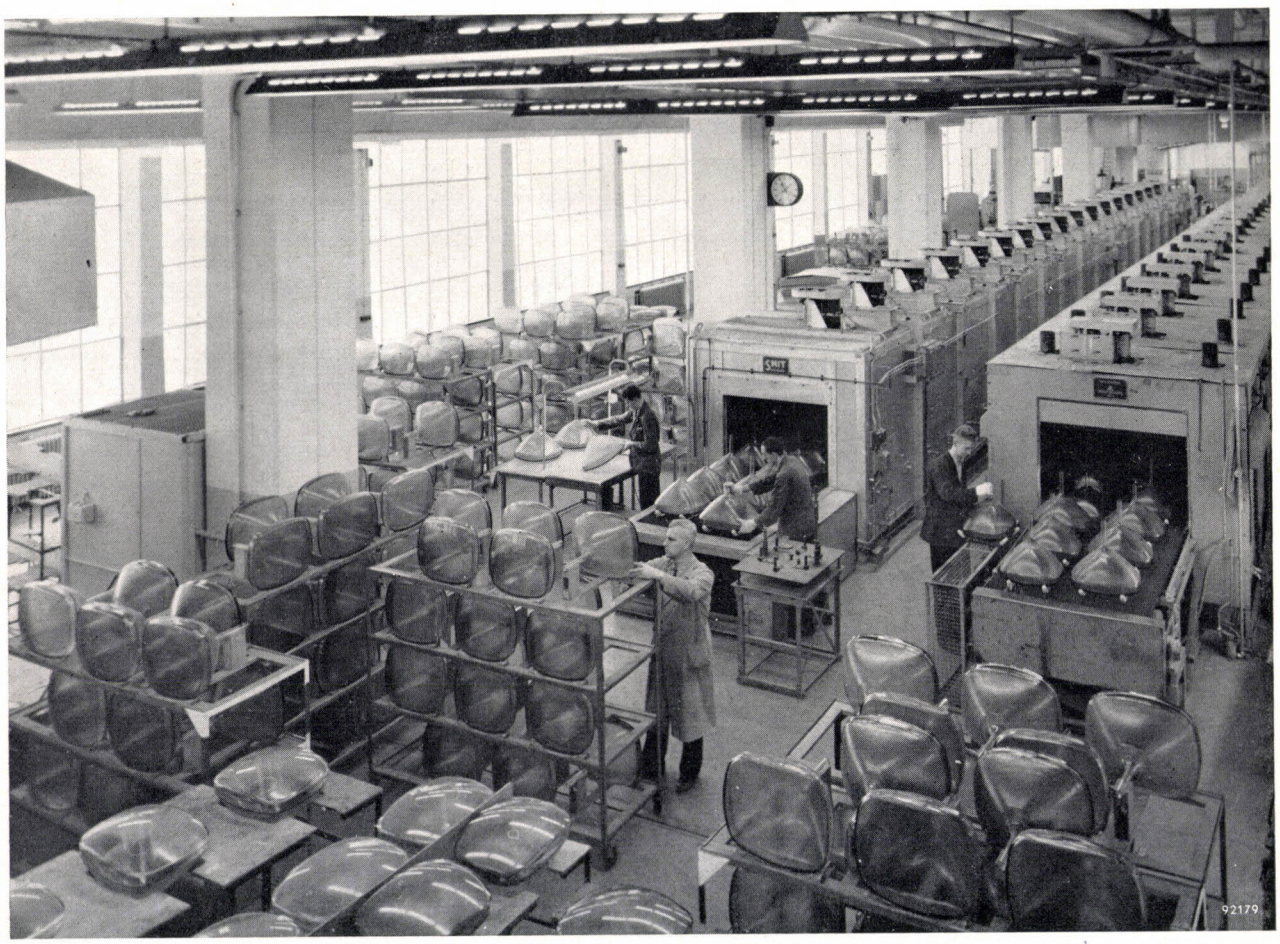
The voltage for the tube is derived from a three-phase high-tension transformer and six rectifiers. The transformer and rectifying valves are immersed in a water-cooled oil bath. The voltage for the transformer primary is derived from a variable three-phase transformer. This latter, and the oil tank just referred to, are separate from the main apparatus and are not visible in the photographs.

The control unit is mounted on a mobile rack. Also on this rack are three smoothing coils forming part of the high-tension circuit (a coil for each phase), and the stabilizer for the filament current. A cut-out, also mounted on the rack, switches off the high-tension supply should the anode current rise above a certain value.

The X-ray diffraction tube described here has now been in satisfactory operation for some years at the Laboratory of Crystal Chemistry of Utrecht University.

W. J. H. BEEKMAN, A. VERHOEFF and H. W. van der VOORN.

ANNEALING THE ENVELOPES OF TELEVISION PICTURE-TUBES



The envelope of a television picture-tube consists of three parts: the screen, "cone" and neck. When these parts are fused together, stresses are set up in the glass. To relieve these stresses, the envelopes are conveyed on a moving belt through a long annealing oven. The above photograph shows the process in operation in the new television tube factory in Eindhoven.

A METHOD OF SEALING THE WINDOW AND CONE OF TELEVISION PICTURE-TUBES

by A. H. EDENS.

666.1.037.4:621.385.832:621.397.62

This article gives some idea of the various problems which arise in the mechanized working of mass-produced heavy glass parts which do not possess rotational symmetry.

Thin-walled glass parts are generally sealed together by heating them with a number of gas jets which are aligned radially, as shown in *fig. 1a*. The hot edges are pressed together, producing a joint as illustrated in fig. 1*b*. By immediately stretching slightly the joined parts, and sometimes rolling the area of the seal, inside and outside, a joint is obtained which, as regards wall-thickness, differs little from the rest of the structure (fig. 1*c*). A case in which such a technique is employed, viz. the sealing of the window and cone of a special type of cathode-ray tube, was described about a year ago in this Review [1]).

At Eindhoven, the above-mentioned difficulty is avoided by using a method of sealing in which the edges to be sealed are heated "head-on", that is to say with gas jets head-on to the edges instead of radially directed jets. The geometry of the method is illustrated in *fig. 2*. We shall now briefly consider some of the special features of this method.

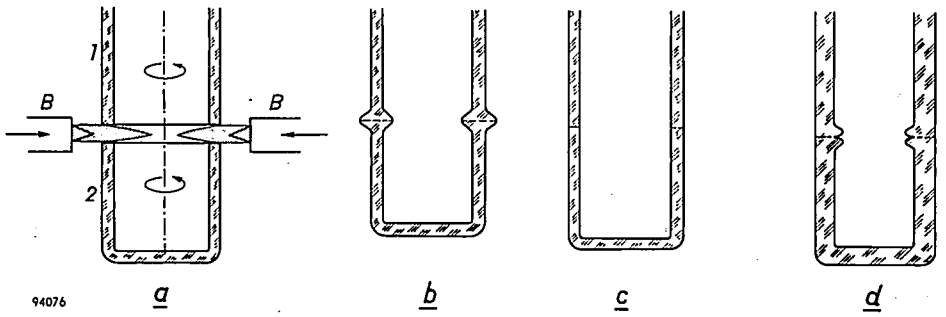


Fig. 1. *a*) For sealing two thin-walled sections *1* and *2* the sections are heated, while rotating, by burners *B*, the jets being directed radially on the glass parts to be joined.
b) Joint obtained by pressing the hot ends together.
c) Joint after stretching.
d) The sealing of two thick-walled sections by the same method does not produce a good joint (notch on the inside).

A satisfactory joint cannot be obtained, however, by using this same method for sealing together thick-walled parts, such as the window and cone of a glass television picture-tube (wall thickness 5 to 6 mm or more). In this case, sealing no longer takes place over the whole surface of the edges, since the edges during heating exhibit a radial temperature gradient (the direction of the gas jets). If the temperature on the outside of the bulb is correct, the temperature on the inside is too low, and consequently the surface tension cannot overcome the high viscosity of the glass at this point. This gives rise to an imperfect joint on the inside (fig. 1*d*). This is particularly undesirable if the joint is subsequently to be subjected to a heavy mechanical load, as it is in the case of an evacuated picture-tube.

During the sealing process the two glass parts are kept rotating, both in the conventional method with radial flames and in the head-on system. If the sections were not rotated, then, in spite of the distribution of the heat supply among many small

[1]) F. G. Blackler, Rectangular cathode-ray tubes of high aspect ratio, Philips tech. Rev. **18**, 298-300, 1956/57 (No. 10).

gas jets, each separate jet would inevitably cause local overheating, giving rise to drops of molten glass and possibly to gas bubbles. Moreover, rotation is desirable to compensate for minor differences

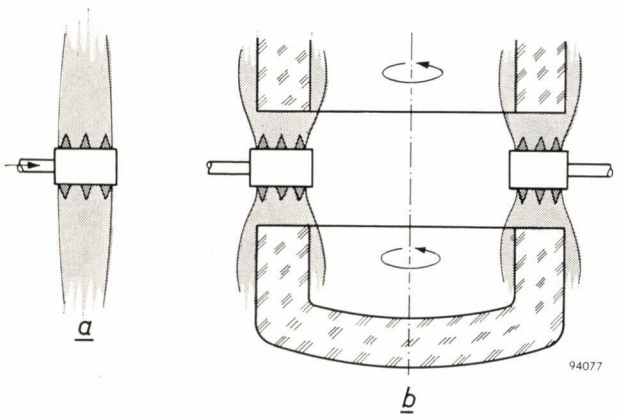


Fig. 2. By the method developed at Philips, thick-walled glass parts are sealed by means of burners (*a*) whose gas jets are directed head-on to the edges to be joined (*b*).

in the functioning of the usual gas-oxygen burners, caused by dirt or slight differences of adjustment: with rotation the heat from each burner is spread out over the entire surface.

For glass parts having rotational symmetry, no problems are entailed by rotation which, depending on the dimensions, takes place at a speed from 15 to 30 r.p.m. The burners are kept stationary with respect to the two synchronously rotating edges, resulting in a uniform relative movement of each gas jet over the whole area of the edge. With picture-tubes, however, the window and cone of which are substantially rectangular with rounded corners, special measures are called for in order to ensure that, during the rotation of the glass parts, each part of the edge receives an equal supply of heat.

In the first place the burners must be moved radially to and fro, so as to keep them always directed on the edges being heated. This requirement is also imposed in the conventional method of radial heating (see the article quoted under [1])). In the head-on method a further requirement is that the box-shaped burner heads, each of which contains a series of jets, above and below, should be swivelled during motion in such a way that the rows of jets are at each instant properly aligned between the edges of window and cone. This requirement is fulfilled with the aid of the mechanism illustrated in *fig. 3*. Each burner is fixed to a small carriage whose two rollers rest against the periphery of a flange which is profiled to the same form as the edges to be sealed. The carriage is pivoted on a block sliding in a radial guide, the block being kept pressed inwards by means of a piston actuated by compressed-air. In this way, then, the burners automatically move radially and swivel so as to maintain correct alignment on the edges. This is not enough, however. A glance at fig. 3 shows that during rotation the four corners of the parts to be sealed together are exposed to each burner longer than the centres of the sides. To ensure uniform heating a further measure is therefore necessary: the fuel feed to the burners must be varied in a specific manner during their movement.

Before we go on to discuss how this is done, we may complete the general description of the process of sealing glass parts by this method. Around the bulb twelve burners are disposed, uniformly positioned over an arc of 247.5° (angular separation 22.5°); see *fig. 4*. As soon as the two edges are heated to a sufficiently high temperature, i.e. after head-on heating for about 15 seconds, the burners are withdrawn from between the edges, and the window section of the picture-tube is pressed against the cone section by an upward movement of the window support. The butt joint thus obtained, and illus-

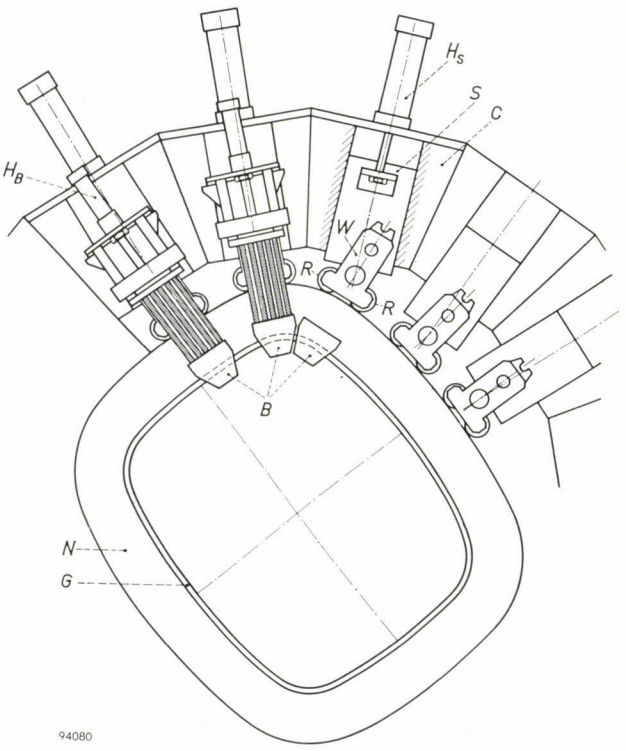


Fig. 3. Mechanism for guiding the burners in such a way that the rows of gas jets in head-on heating are always directed on the edges G of the substantially rectangular glass sections. Only two of the twelve burner systems are shown complete in the drawing. Each burner head B is mounted on a carriage W, which is pivoted on a block S and which, with two rollers R, is kept pressed against the profiled flange N by a compressed-air actuator H_S (C is a guide for the block S). Each burner can be moved forward (for head-on heating) by a small compressed-air actuator H_B, mounted on the carriage, and drawn back for after-heating.

trated in fig. 1*b*, is stretched by a subsequent lowering of the window support by a few millimetres: the joint then stretches by virtue of the weight of the window. In order to prevent the temperature of the joint from dropping too much during the time needed for the stretching (or "sagging") process, after-heating is applied by means of radially directed gas jets. The manner of stretching can be controlled by applying the after-heating higher or lower in the area of the seal, and in this way joints of good uniformity in wall thickness can be obtained.

The after-heating, during which the bulb continues to rotate, is effected with the same burner heads by flames from rows of jets in the front face of the heads; see *fig. 5a* and *b*). These jets and their fuel feed constitute an entirely separate system. When the burners are withdrawn from between the edges, the fuel feed is automatically switched over from the system for head-on heating to the system for after-heating. The withdrawal of the burners is brought about by separate compressed-air actuators which move the burners with respect to the swivel-

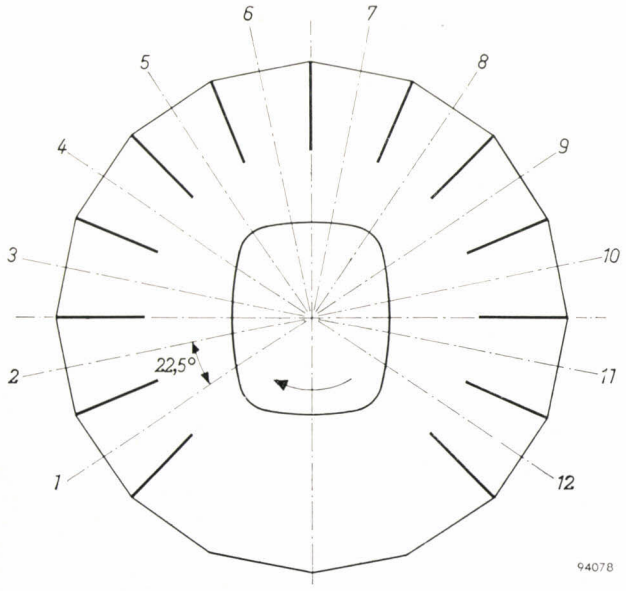


Fig. 4. Arrangement of the burners *1* to *12* around the glass parts to be sealed. The workpieces are accessible from the front.

ling, rollered carriages, the rollers always remaining pressed against the profiled flange. Thus the burners continue their radial and swivelling motions during the after-heating of the rotating bulb. For after-heating, the same periodic regulation of the fuel feed is necessary as for head-on heating.

The periodic variation required for the fuel feed can be derived from *fig. 6a*. In this figure, which refers to a 17″ picture-tube, the relative speed of

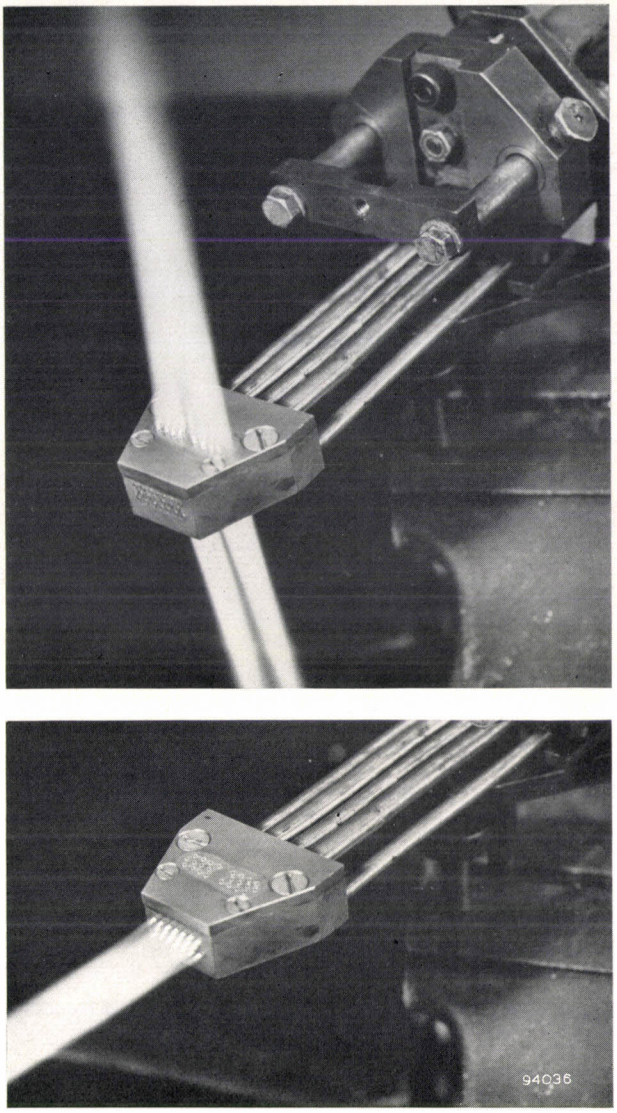


Fig. 5. The burner head, (*a*) with the jets for head-on heating in operation and (*b*) with the jets for after-heating in operation.

a burner with respect to the glass passing under it is plotted against time, or rather — which amounts to the same thing — against the angular position α of the radius vector of the carriage (fig. 6*b*). Without the swivelling of the burner the result would be the smaller variation indicated by the dashed curve in fig. 6*a*; each maximum of this curve corresponds to the passing of one corner of the window, where the linear speed of the glass is, of course, greatest. However, since the swivelling of the burner carriage, which at that point is opposed to the linear motion of the glass, is also fastest at the corners, a deep minimum is in fact produced at this point in the relative speed of the glass with respect to the burners; see the fully-drawn curve. It can be seen from the figure that the relative speed during the rotation of the glass varies by about a factor of 2.

Since slight temperature differences along the edges to be joined can be tolerated, it is not necessary that the variation of the fuel feed should follow accurately the form of the curve in fig. 6*a*. In practice a periodic variation between two levels is found to be sufficient (except for a small modification to be mentioned later). This variation is carried out in the following way.

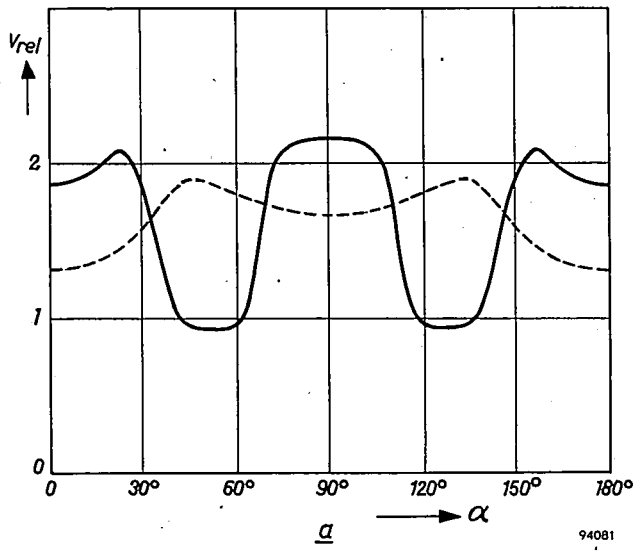


Fig. 6. *a*) Relative speed (v_{rel}) of a burner with respect to the glass, as a function of the position (angle α) of the burner carriage with respect to the revolving glass parts. The broken curve is that which would be obtained if the burner, instead of swivelling, were always pointed radially inward. *b*) Sketch illustrating the definition of α.

The two systems (for head-on and after-heating) in each burner are each fed via a separate ejector device to which oxygen is supplied under pressure and which sucks in gas. Ejector-fed burners of this kind generally operate on an excess of gas, as they do in our case. The primary flame, the familiar blue cone, forms directly over the holes of the burner; the amount of gas burnt here corresponds to the quantity of oxygen in the gas-oxygen mixture. The excess gas burns on the air in the atmosphere. Since each burner head with its jets moves at a distance of only about 7 mm from the edges of the glass parts, the quantity of heat delivered at any instant to these edges is mainly determined by the flow of oxygen at that moment from the burner. The burner is therefore regulated by varying the pressure of the oxygen fed to the ejector. The regulating device is represented in *fig.* 7. The oxygen flows through the device via two openings, a wide one which can be closed by a valve, and a much narrower one (a constriction). The valve is actuated by a cam which releases the valve stem (this being under spring pressure), thereby closing the valve whenever one of the four corners of the glass parts approaches the particular burner. The oxygen can then flow only through the constriction, over which an appreciable pressure drop exists; the open valve, on the other hand, causes almost no pressure drop. The oxygen pressure at the ejector inlet thus varies periodically between a low value (at the corners) and a high one.

We shall now mention some details of the actual mechanism. The regulator illustrated in fig. 7 supplies oxygen to the ejectors of two diametrically opposite burners — for example No. 1 and No. 9 in fig. 4 — since two oppositely situated burners must receive the same oxygen sypply at each instant. The four burners Nos. 5 to 8, which have no opposite numbers, each have a regulator of their own. For head-on heating there are therefore eight regulators, which are incorporated in a single block and whose valves are actuated with appropriate phase differences by eight cams mounted on one rotating shaft. This camshaft must of course have a certain fixed angular adjustment with respect to the profiled flange around which the burners move (and hence with respect to the revolving glass).

For the after-heating systems the regulation is carried out in an identical manner; here too there

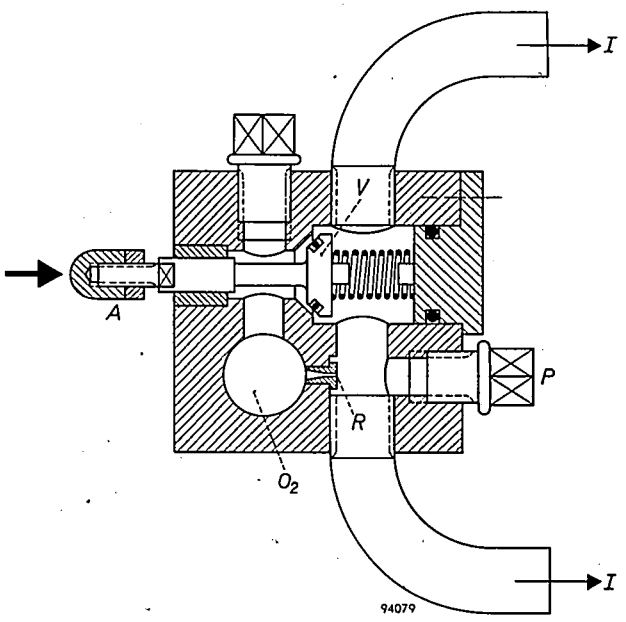


Fig. 7. Regulator for periodically varying the pressure under which the oxygen is fed to the ejector *I* of a burner. (This regulator serves two ejectors.) O_2 oxygen feed to regulator; *V* valve, which is opened by the action of a cam on the stem (adjustable with nut *A*) against spring pressure; *R* constriction, the only route which the oxygen can take when *V* is closed, resulting in a reduced pressure in the pipe to the ejector(s).

are four double and four single regulators. These are incorporated in a second block, in the same order as the regulators for head-on heating, so that they can be operated by the same cams. The two series of regulators differ only in the size of the constrictions, since the burner jets for head-on heating pass a greater quantity of fuel mixture than those for after-heating. The constrictions in the regulators with a double ejector connection are, of course, also wider than those with a single connection. The size of the constrictions is chosen according to the bulb under production (that is to say, to the curve in fig. 6*a*). For the production of other bulbs, larger picture-tubes for example, the constriction in each regulator can be changed after first removing the plug P in fig. 7. It is also necessary, as a rule, to change the cams.

In one full revolution of the glass each cam must complete two identical cycles. For simplicity, therefore, the cams have only two flats instead of four (for the four corners), and the cams perform two revolutions for each revolution of the glass workpiece. With a view to convenience of installation and easy access to the supply pipes, the two blocks containing the regulators for head-on and after-heating (K' and K'' respectively, *fig. 8*) are mounted directly facing each other at opposite sides of the cams. This entails that regulators K' and K'', which are actuated by the same cam, must operate on burners which are mounted in the machine at an angular separation of 90°. Thus the cam which actuates regulator $K'_{1,9}$ is also responsible for K''_5, the cam for $K'_{2,10}$ also relates to K''_6, and so on. In this way, in after-heating as well as in head-on heating, all burners are automatically controlled in the correct phase with respect to the revolving glass parts.

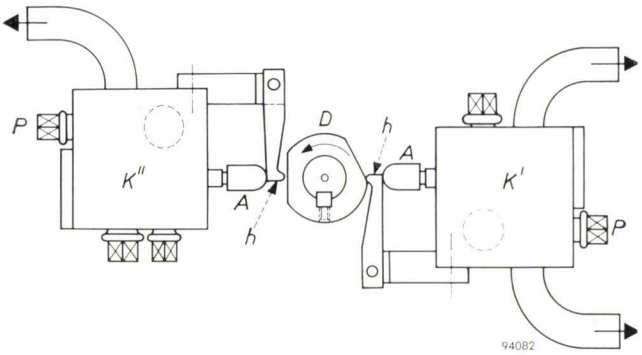


Fig. 8. The block containing eight adjacent regulators K' for head-on heating, and a similar block with regulators K'' for after-heating, are both driven by a single camshaft with eight cams D. The end A of each valve stem is not directly in contact with its associated cam, but rests against a separate rocker arm h, so that the valve stem is subjected only to an axial load.

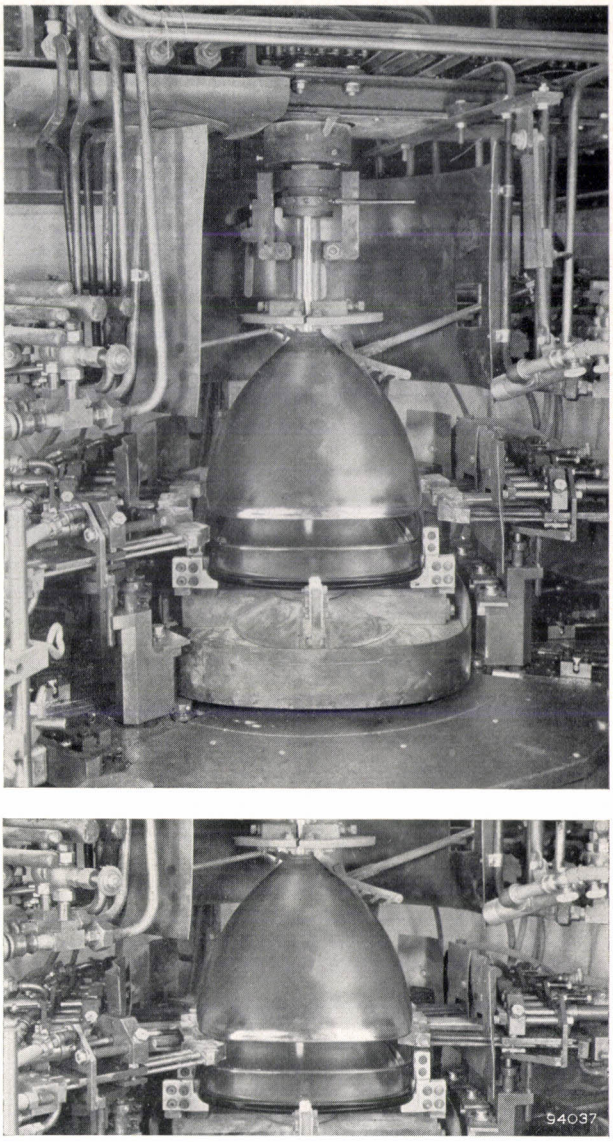


Fig. 9. *a*) View of the glass-sealing machine, showing the window and cone sections of a 17″ picture-tube in position. The burners are withdrawn.
b) Burners moved forward between the edges to be sealed (in the actual process the burners are not moved into position until the glass has first been pre-heated for a certain period; to avoid photographic complications, the pre-heating stage was omitted in taking these pictures).

In the set-up as described, all twelve burners pass through the same control cycle. In principle this is not necessary, since at each point of the edge the glass will experience the average effect of all burners. In fact, in the case of one pair of burners for head-on heating, a modification is made to the cycle described: the cam in this case is provided with an additional flat such that the burners in question (and the after-heating burner served by the same cam) receive less fuel in the middle of the long side of the window and cone. Before this modification was applied a hot spot in the glass was found corresponding to the trough at $\alpha = 0°$ in the curve in fig. 6*a*.

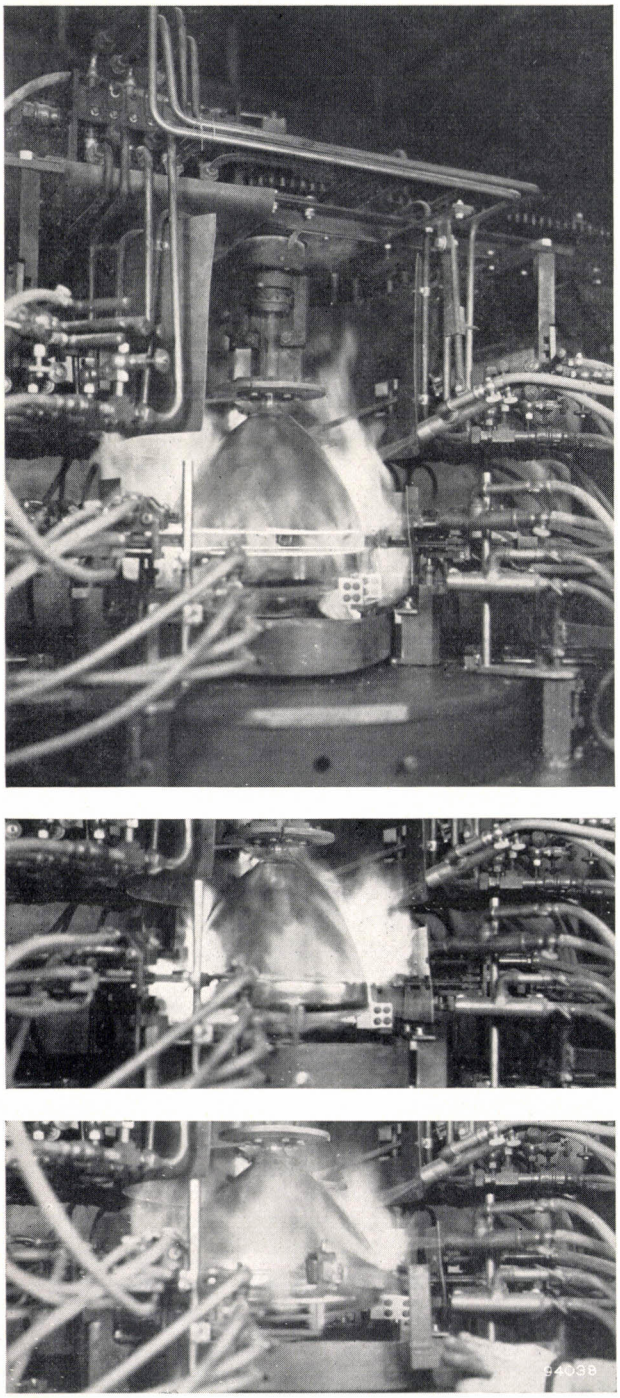


Fig. 10. A tube of the type shown in fig. 9 in the process of being sealed:
a) The edges to be joined are heated to the required temperature.
b) The edges have been pressed together and after-heating has begun, while the joint is allowed to stretch.
c) Rolling of the joint after stretching.

Finally, we give a number of photographs of the glass-sealing machine, which illustrate the manufacture of a 17″ picture-tube. To facilitate the handling of the large glass components, both before and after sealing, the machine is constructed such that the picture-tube is mounted in it with its axis vertical. *Fig. 9a* shows the window and cone section mounted in position and with the burners withdrawn; in fig. 9*b* the burners are seen in the sealing position, the burner heads being between the edges of window and cone.

After the sections have been set in rotation and preheated, the head-on heating begins; this lasts, as stated earlier, about 15 seconds, during which time the glass completes about four revolutions. In *fig. 10a*, head-on heating has just been completed and the burners are again in the withdrawn position. The edges of window and cone are pressed together, after-heating of the sealed zone begins (fig. 10*b*) and the window support drops slightly to stretch the joint. One detail worth mentioning is that at the moment the gap between window and cone is sealed, a slight, adjustable over-pressure is created inside the bulb. This serves to compensate for the pressure exerted by the gas jets on the area of the seal during after-heating, and which might otherwise cause this area to collapse. After the joint has been stretched, it is smoothed on the outside with a graphite roller (fig. 10*c*).

Most of the operations described are performed automatically, being programmed by means of a camshaft at the rear of the machine.

Summary. For sealing thick-walled glass parts, such as the window and cone of a picture-tube, a method is used at Philips in which the glass parts, while rotating, are heated by gas-oxygen jets directed head-on to the edges to be sealed. Twelve burner heads are placed between the edges of the roughly rectangular workpieces, which are in rotation; by means of a guide mechanism the heads are simultaneously moved radially and swivelled in such a way that the rows of burner jets are always directed on the edges of the workpieces. To compensate for the periodic variations in the speed of each burner relative to the glass, the fuel feed (oxygen pressure) of each burner is varied periodically in the correct phase. This also applies to the after-heating of the seal by the same burner heads.

FABRICATION OF SAPPHIRE AND DIAMOND STYLI FOR GRAMOPHONE PICK-UPS

681.84.081.32

Gramophone pick-ups with sapphire and diamond styli have largely replaced the older pick-ups which used interchangeable steel needles. One important reason for this lies in the enormously lower wear suffered by these hard materials. Although they are very much harder, sapphire and diamond cause much less record wear than steel needles: it is found that both distortion of reproduction and record wear are more dependent on the *shape* of the needle point than on other factors, so that the longer this is maintained the better. The photographs of *fig. 1* show how well sapphire and diamond needles maintain their shape, even after hundreds of hours of use. A steel needle wears to the groove shape after only one playing. Another important advantage of the sapphire and diamond styli is that the moving system of the pick-up can be made with a much smaller moment of inertia than in the older types with their comparatively massive steel needle and holder. In modern pick-ups using sapphire and diamond styli, the effective mass at the needle point may be as little as 3 mg, resulting in a cut-off frequency due to stylus/groove resonance of the order of 20 kc/s, which is well above the limit of the audible frequency range [1]).

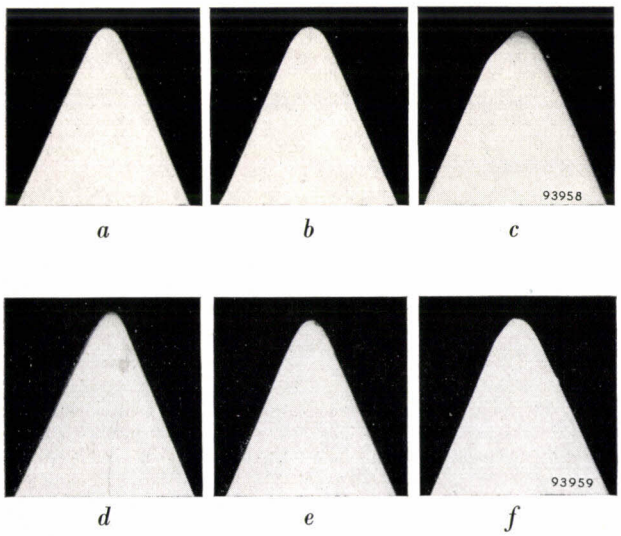


Fig. 1. Comparative shadow photographs showing the wear of a microgroove sapphire stylus and a microgroove diamond stylus (tip radius 25 μ). The vertical force on the needle amounted to 12 gram. For each stylus a new long-playing vinylite record was taken and the test continued throughout with the *same* record. It should be stressed that under these conditions the wear on the needle is far higher than ever occurs in practice. The microscopic particles worn from the needle are in normal circumstances distributed over a whole collection of records; here they are concentrated in the grooves of a single record, resulting in a much higher rate of wear. Furthermore, the photographs give no quantitative indication of the advantage of the diamond over the sapphire; in fact, the diamond may have a life of more than twenty times that of a sapphire point. *a*) Unused sapphire stylus. *b*) Sapphire, after 75 hours. *c*) Sapphire, after 175 hours. *d*) Unused diamond stylus. *e*) Diamond, after 150 hours. *f*) Diamond, after 300 hours.

Sapphire and diamond styli for gramophone pick-ups are manufactured in the Philips diamond-die factory at Valkenswaard (near Eindhoven).

Sapphire cylinders 3 mm long and 0.4 mm in diameter, cut from sheets of synthetic sapphire, are mounted in small rotating heads by means of shellac. A number of such heads are placed at an angle ($22\frac{1}{2}°$) to an equal number of cast-iron grinding wheels, the pores of which contain a little diamond dust in oil. The heads and the grinding wheels are rotated at high speed while at the same time the grinding wheels are given a slow reciprocating movement parallel to their surface in order to obtain uniform grinding (*fig. 2*).

In this way the cylinder is given a conical end. The process is stopped, however, before the cone is sharp-pointed, as this tends to cause breakages during further handling. The sapphire cylinder is then reversed and the coning operation repeated on the other end. Subsequently the needles are broken in half, ground to a length of 1 mm and cleaned, and are then ready for the radiusing of the points. This is done by placing a large number of them (several thousands) together in a bottle containing diamond dust in oil. The bottle is then vibrated for several days, after which the points are found to be spherical in shape. The radius becomes greater as the vibration is continued. In this way the required radius (25 μ for microgroove records, 75 μ for normal 78 r.p.m. records) is easily obtained [2]).

For diamond styli, selected stones are first cut parallel to the (111) plane to give plates 1.6 mm in thickness (*fig. 3*). From these plates cylinders 0.3 mm in diameter are cut by means of thin-walled nickel tubes rotating at about 10 000 r.p.m. and containing diamond dust in oil. The ends of the cylinders which will become the points of the styli are therefore in the (111) plane. This plane is chosen partly because of its great wear resistance [3]) but

[1]) For an explanation of these terms and a fuller discussion of the questions involved, see J. B. S. M. Kerstens, Mechanical phenomena in gramophone pick-ups at high audio frequencies, Philips tech. Rev. **18**, 89-97, 1956/57; also N. Wittenberg, A magnetodynamic gramophone pick-up, Philips tech. Rev. **18**, 101-109 and 173-178, 1956/57.

[2]) The same technique is used for rounding off the edges of quartz oscillator crystals.

[3]) See, for example, L. Schultink, H. L. Spier and A. J. van der Wagt, Philips tech. Rev. **16**, 91-97, 1954/55.

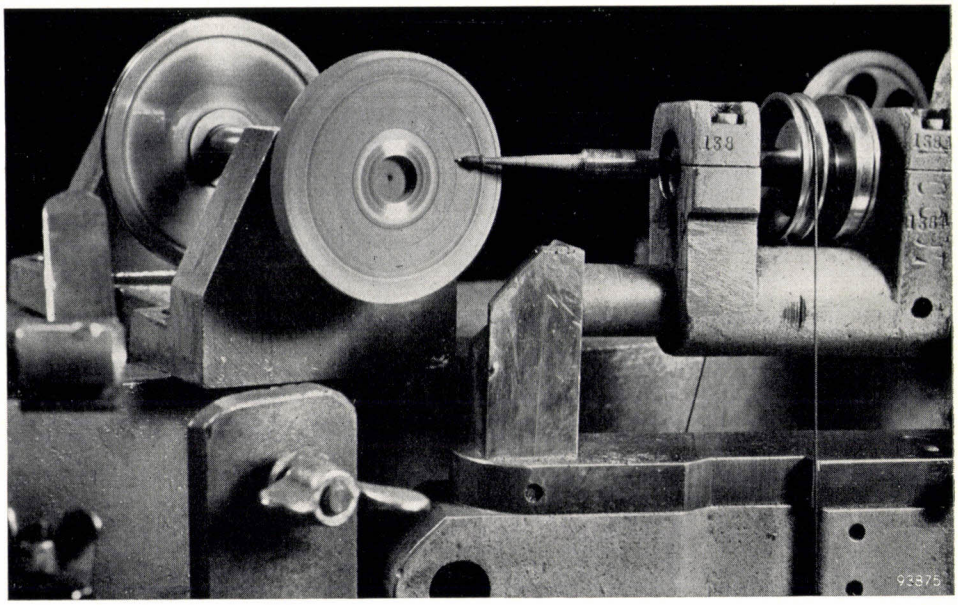


Fig. 2. Part of the 10-head machine for grinding cones on the ends of sapphire and diamond cylinders. The frame of the machine is suspended on four flexible steel strips. At one end of the machine a cam imparts a slow reciprocating movement to the whole frame, parallel to the surfaces of the grinding wheels. The sapphire or diamond cylinders are fixed in the rotating heads by means of shellac.

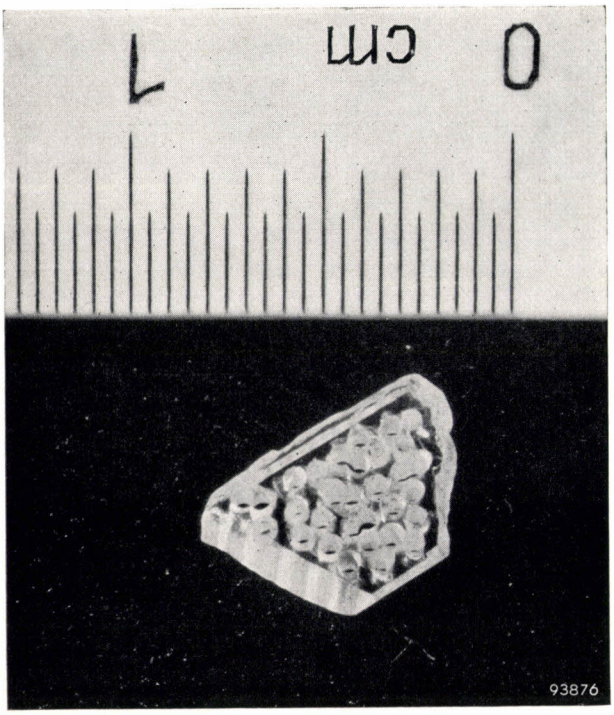


Fig. 3. Diamond platelet, 1.6 mm thick, cut parallel to the octahedral (111) plane and drilled out to give cylinders of 0.3 mm diameter (see fig. 4, above).

also because (111) faces of reasonable area can be cut from the natural crystal without too much loss of material. Next a cone is cut at each end of the cylinder in the same way as for sapphire. In the case of diamond, however, the operation takes about 10 times longer. The radiusing of the ends cannot be effectively performed by the technique used for sapphire in view of the greater hardness and the fact that the mass of the diamond styli is considerably less. Diamond styli are therefore radiused by hand on a special grinding wheel. The rotating head containing the diamond stylus is removed from the coning machine and set in the radiusing machine; the latter is equipped with a microscope for control of the operation. The radiusing takes about one minute. The diamond cylinder is then reversed in the rotating head and replaced in the coning machine; this is again followed by the radiusing operation. After cleaving the pointed cylinder in half and cleaning, the styli are ready for mounting. *Fig. 4* gives an idea of the size of the diamond cylinder and the finished stylus.

Of the original diamond, less than half can be drilled out into cylinders of the required dimensions. The remaining drilled stones are crushed to form diamond dust. Diamond dust is also recovered from

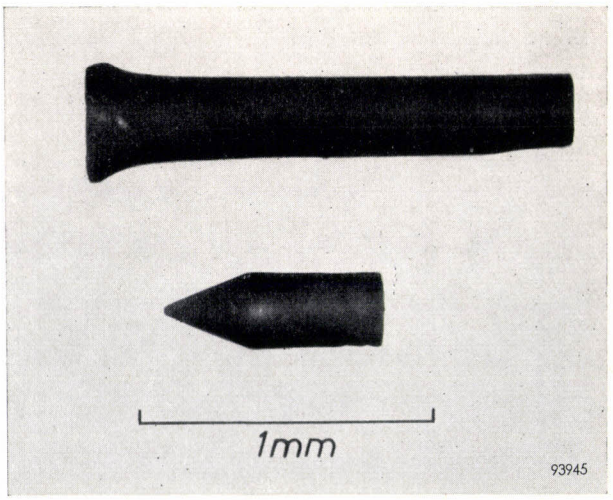


Fig. 4. Above, shadow photograph of diamond cylinder as bored from diamond platelet; below, shadow photograph of a finished stylus, 0.3 mm diam., 0.6 mm long.

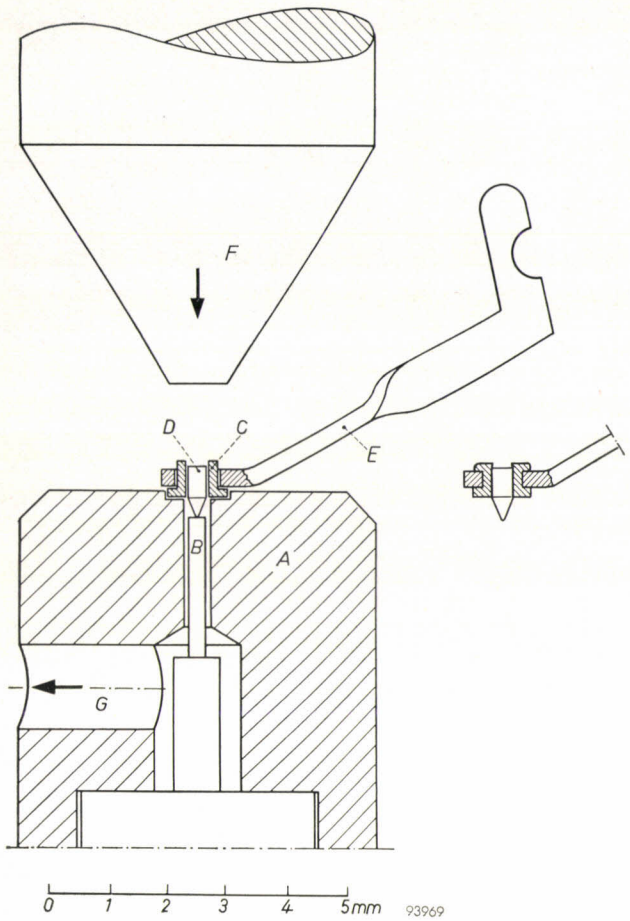


Fig. 5. Sketch of cleating jig (about 8× actual size). *A* vacuum jig, *B* stop, *C* aluminium bush, *D* stylus, *E* stylus arm, *F* punch, *G* vacuum line.

the vicinity of the various grinding operations by means of carefully placed shields. After cleaning and grading into size ranges by centrifuging, the dust can be used again. There is, however, a surplus (mainly of the finest grades) and this is used for other purposes, e.g. for polishing in metallography.

Both the sapphire and the diamond styli are mounted on the stylus arm by the same technique.

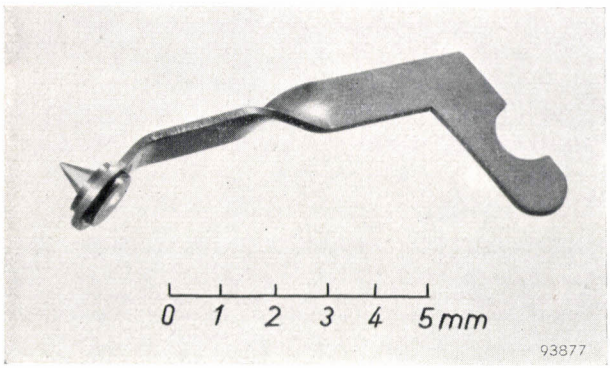


Fig. 6. Finished stylus mounted on stylus arm.

A tiny aluminium bushing is picked up on the end of a 0.3 mm bore nickel tube through which air is sucked (tweezers are not very suitable for handling such small objects). The bushing is placed on a hollow jig mounted in a light press (*fig. 5*). Air is also sucked through this jig and the dimensions and pressure are so chosen that the bushing is taken over from the nickel tube and seated in the correct position. A stylus is then placed inside the bushing, point downwards; the suction causes the point to come up against a stop which ensures a constant projection of the stylus below the stylus arm. The latter is placed round the bushing and the press operated, providing a completely rigid mounting for the stylus (*fig. 6*).

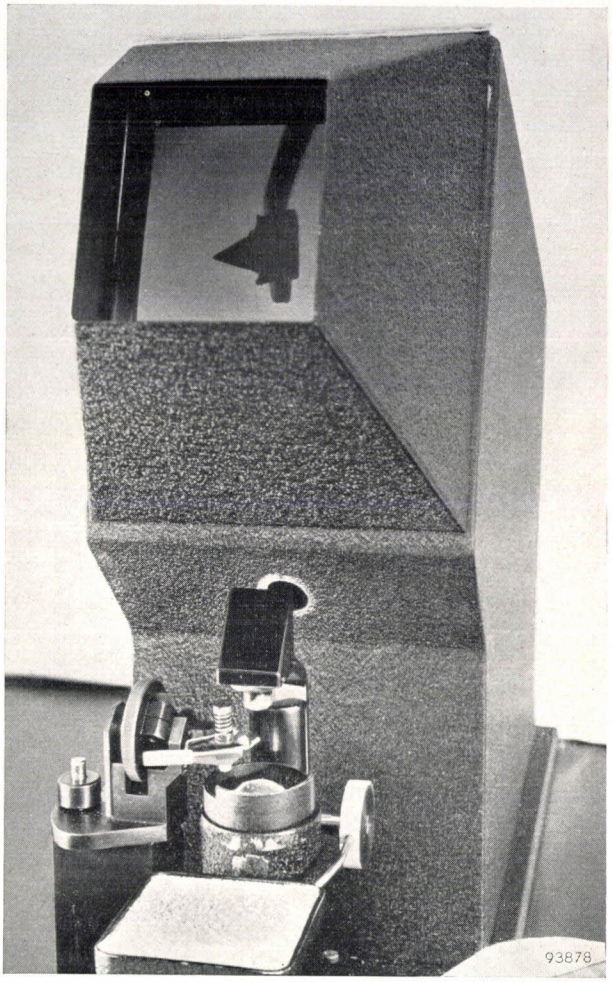


Fig. 7. Projection microscope (magnification 50×) for the individual inspection of mounted sapphire and diamond styli. The three lines on the screen serve for the checking of dimensions and alignment.

The mounted styli are then individually checked in low-power projection microscopes (*fig. 7*).

A. W. PLOEGSMA.

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.

2512*: J. Volger and J. M. Stevels: Continuation des recherches expérimentales sur les pertes diélectriques de certains verres aux basses températures (Verres et Réfractaires **11**, 137-146, 1957, No. 3).

Translation of article in Philips Res. Repts **11**, 452-470, 1956 (see these Abstracts No. **R 301**).

2513: J. H. Uhlenbroek and M. J. Koopmans: Investigations on agricultural fungicides, II. Compounds structurally related to trichloromethyl thiolsulphonates (Rec. Trav. chim. Pays-Bas **76**, 657-665, 1957, No. 7).

The preparation and the fungitoxic properties of some thiolsulphonic esters (type p-$CH_3C_6H_4SO_2SR'$) and of some reaction products of trichloromethanesulphenyl chloride with thiosulphonic acid salts and with thiophenols are described in this paper.

2514: J. H. Uhlenbroek and M. J. Koopmans: Investigations on agricultural fungicides, III. The reaction between trichloromethanesulphenyl chloride and benzoic acid (Rec. Trav. chim. Pays-Bas **76**, 666-668, 1957, No. 7).

A mixed anhydride of trichloromethanesulphenic acid and benzoic acid has been prepared for which the name of benzoic trichloromethanesulphenic anhydride is proposed. Some chemical and biological properties of the new compound are described.

2515*: K. Compaan and Y. Haven: La diffusion de radiotraceurs dans les solides;
J. L. Meijering: Diffusion des atomes en insertion;
A. van Wieringen: Etude du passage non-stationnaire de l'hélium à travers le silicium et le germanium à l'aide d'un spectromètre de masse
(La diffusion dans les métaux, edited by J. D. Fast, H. G. van Bueren and J. Philibert, Philips Technical Library, 1957).

See book notice below, p. 332.

2516: S. Duinker: L'application des ferrites à la modulation dans divers domaines de fréquence (Atti del Congresso Scientifico, sezione elettronica, July 1956, pp. 571-585).

Survey of ferrite applications for modulation in tuned circuits, in phase and frequency modulators, in magnetic modulators and in absorption modulators for S.H.F.

2517: E. G. Dorgelo: Some technological aspects of U.H.F. triode design (Le Vide **12**, No. 67, 3-8, 1957).

It is well known that the efficiency of a triode decreases at higher frequencies. This is due to the fact that the transit time of the electrons is no longer negligible in comparison with the period of the oscillations. The transit time can be made small by mounting the electrodes very close together and by applying large potential differences. Both measures, however, give rise to large current densities. As a result, the design of high-frequency transmitting valves comes against a number of technological problems, of which two are mentioned in this article: cathodes with a high current-density emission and grids with a large specific dissipation. A new material for the grids (K-material) is described for which continuous dissipations of up to 30 W per cm^2 are possible. The high-frequency resistance is sufficiently low to permit use in transmitting valves for decimetre waves. The article concludes with a short description of such a valve.

2518: E. G. Dorgelo: Gitterprobleme bei Sendetrioden in Schaltungen der industriellen Elektronik (Funk-Technik **12**, 528, 530 and 532, 1957, No. 15). (Grid problems in transmitting triodes used in industrial electronics; in German.)

It is common practice to include in the published data of transmitting valves, so-called limiting values for grid input. As long as this value is not exceeded, there ought to be no excessive thermal emission from the grid. There are three contributions to the total power, dissipated in the grid: heat generated by electron bombardment, absorbed radiation from other electrodes and R.F. losses. In general, the user can measure only the first mentioned. Even here errors can be made due to the fact that the measured grid current is not representative of the number of electrons arriving at the grid. When secondary emission is present, other electrons will leave the grid and the measured current is smaller than the actual grid current. In view of t

fact that the valve manufacturer usually does not give an indication what he actually means by the published value and also because it is physically impossible to measure the true grid currents, it is rather difficult for the user to evaluate the quality of a valve merely from the published data. In an example the author compares two equivalent valve types of different make. Measurements reveal that the valve with the best-looking published data shows the highest thermal emission of the grid.

2519: H. P. J. Wijn, H. van der Heide and J. F. Fast: Ordering in cobalt-ferrous ferrites (Proc. Instn. Electr. Engrs. **104 B**, suppl. No. 7, 412-417, 1957).

Various cobalt-ferrous ferrites show a constricted hysteresis loop. After magnetic annealing of the samples, the loop becomes rectangular. It appears that the magnetic annealing creates in each crystal a uniaxial anisotropy in a direction which is not necessarily the direction of the applied magnetic field but a crystallographic direction nearest to it. It is suggested that directional ordering is the most probable origin of the anisotropy found.

2520: E. W. Gorter and C. J. Esveldt: Square-loop ferrites obtained by magnetic annealing of new compositions (Proc. Instn. Electr. Engrs. **104 B**, suppl. No. 7, 418-421, 1957).

Extension of the work abstracted in No. **2519** to cobalt ferrites of other composition. When a ferrite containing cobalt is cooled in a magnetic field, diffusion of Co^{2+} ions causes a uniaxial anisotropy that is apparent in a rectangular hysteresis loop. In various ferrites containing only a small proportion of Co^{2+} ions, this diffusion is very slow. For the composition $(Mg_{0.6}Ni_{0.4})_{0.98}Co_{0.02}Fe_2O_4$, however, this diffusion is fast; on cooling in a magnetic field, a weak coercive force and an extremely rectangular hysteresis loop are obtained. The loop is practically independent of temperature.

2521: H. P. J. Wijn and H. van der Heide: Pulse-response properties of rectangular-loop ferrites (Proc. Instn. Electr. Engrs. **104 B**, suppl. No. 7, 422-427, 1957).

An analysis of the voltage pulses obtained across a secondary winding of a ferrite core on reversing the magnetization shows that in the reversal two stages should be distinguished, namely a fast response, related to a magnetization by rotations, and a slower response due to the displacements of domain walls. The viscous movement of the domain walls determines the switching time of the core. It appears that the switching times of rectangular-loop ferrites with widely varying chemical compositions are always of the order of magnitude of one microsecond when the coercive force is of the order of magnitude of one oersted. Higher values of the coercive force mostly go hand in hand with an increased switching time. In order to obtain some information about the origin of the damping of domain-wall displacements, switching times have been measured as a function of temperature and as a function of a uniaxial pressure applied to the core. In the temperature range —115 to +250 °C the switching time of an arbitrarily chosen rectangular-loop ferrite did not decrease more than by a factor of about three. A uniaxial pressure applied to a ferrite also increases both its coercive force and its switching time. It appears that the switching time is greater when the anisotropies in the ferrite are greater. This can at least partly be explained by the increased distance between the walls in the case of higher magnetic anisotropies.

2522: C. M. van der Burgt: Ferrites for magnetic and piezomagnetic filter elements with temperature-independent permeability and elasticity (Proc. Instn. Electr. Engrs. **104 B**, suppl. No. 7, 550-557, 1957).

In recent experiments the compositions and methods of preparation of various ferrites were varied with the aim of achieving optimum magnetic, magnetomechanical and mechanical performance. Essential improvements of the temperature dependence of the permeability, the piezomagnetic coupling coefficient and the mechanical resonant frequencies have been obtained. These were achieved by small cobalt substitutions in nickel and lithium ferrite and in mixed nickel-zinc and lithium-zinc ferrites. The total variations of the mechanical resonant and anti-resonant frequencies of ferrite filter elements in the temperature range 20-50 °C are normally 0.10-0.25% in existing commercial nickel and nickel-zinc ferrites, but have been reduced to 0.03% or less. The coupling coefficient at remanence of such a stable ferrite may be well above 0.20 with a suitably modified sintering treatment. Since the mechanical Q-factors are usually much better than 2000, such ferrite vibrators can be profitably applied to the construction of electrical and electromechanical band-pass filters. The cobalt substitutions, apart from leading to a very low temperature coefficient of the real part of the permeability, also decrease the imaginary part, so that these cobalt-substituted ferrites have a high figure of merit. (See also Philips tech. Rev. **18**, 285-298, 1956/57 (No. 10), and No. **R 310** of these Abstracts.)

2523: E. W. Gorter: Chemistry and magnetic properties of some ferrimagnetic oxides like those occurring in nature (Adv. Phys. **6**, 336-361, 1957; No. 23).

This review paper, given for the Conference on Rock Magnetism (London, November 1956), mainly deals with previous work on ferrimagnetic oxides with spinel structure, described in Nos. **R 248, R 249** and **R 253** of these Abstracts, inasfar as it has a bearing on the ferrimagnetic oxides occurring in nature. In § 1 a more suitable representation of the composition diagram FeO-Fe_2O_3-TiO_2 is proposed. § 2 treats the cation distribution among tetrahedral and octahedral lattice sites and its methods of measurement, the valency of the transition ions present and the saturation magnetization as depending on composition, and discusses the difficulty of establishing the presence of miscibility gaps in a solid-solution series. § 3 discusses the relative orientation of the ionic magnetic moments in the spinel, hematite and pseudobrookite structures, as derived from Anderson's theory. It is shown that none of the explanations so far given for the behaviour of ilmenite, $FeTiO_3$, are satisfactory. In § 4 attention is drawn to the probability of exchange coupling between two phases as a governing mechanism for reverse thermoremanent magnetization.

2524: J. L. Meijering: Diffusion du fluor dans l'argent solide (Rev. Métallurgie **54**, 520-522, 1957, No. 7).

By heating silver containing e.g. 0.05% Mg in molten AgF, two diffusion zones are obtained. The two sharp boundaries are not affected perceptibly by grain boundaries. This "internal fluorination" causes hardening. Probably the inner zone contains MgF_2 and the outer zone a double fluoride like $AgMgF_3$.

2525: F. C. de Ronde: Un élément simple pour mesurer des impédances en ondes centimétriques et millimétriques: La terminaison variable à réglages indépendants et à lecture directe du module et de l'argument du coefficient de réflexion (Arch. des Sci. **10**, fasc. spéc. 6e Colloque Ampère, Rennes-St.-Malo, April 1957, pp. 66-67).

Brief note on a calibrated, direct-reading microwave impedance. An absorbing vane is fixed on a plunger which can be moved axially and also rotated over 90° in a circular waveguide. The modulus of the reflection coefficient depends on the angular position of the plunger and its argument on the axial position.

2526: F. C. de Ronde: Une nouvelle méthode de mesure de la constante diélectrique et de la perméabilité magnétique des matières solides en ondes centimétriques (Arch. des Sci. **10**, fasc. spéc. 6e Colloque Ampère, Rennes-St.-Malo, April 1957, pp. 68-70).

Brief description of a new method for measuring dielectric constant and magnetic permeability in solids at microwave frequencies. The method involves direct measurement of the characteristic impedance of a waveguide filled with the material.

2527: K. Rodenhuis: A 4000 Mc/s triode with L-cathode construction and circuit (Le Vide **12**, No. 67, 23-31, 1957).

Two valves, the EC 56 and the EC 57, for use in beam transmitters are described in this article. The use of L-cathodes, which give long life even at high current densities, and of discs of high H.F. conductivity, leads to an efficiency and a gain which are relatively high for triodes. Some details of the design and of the method of manufacture are given. The behaviour of the valves in an amplifier is discussed. The input impedance is given particular attention. The feedback is found to depend closely on the amplification factor of the valve and to exert an influence on the output power attainable. Finally various other applications of the valves are mentioned. (See also Philips tech. Rev. **18**, 317-324, 1956/57 (No. 11) and **19**, 145-156, 1957/58 (No. 5).)

2528: J. Verweel: The space-charge distribution in a static magnetron (Le Vide **12**, No. 67, 32-42, 1957).

In a non-oscillating cut-off magnetron the motion of the electrons gives rise to a space-charge cloud round the cathode. It is shown that by means of a narrow electron beam, injected parallel to the cathode, the electron paths can be imaged on a fluorescent screen. The electrons are found to describe cycloidal paths starting from the cathode; the radial velocity can be derived from the tangential component of the (thermal) initial velocity. The paths of the electrons are found to be strongly dependent on their initial velocity. To a first approximation, the density of the space-charge cloud is constant.

2529: A. Versnel: Magnetless magnetron (Le Vide **12**, No. 67, 59-63, 1957).

An oscillator tube is described which shows some resemblance to a magnetron both in construction and in operation. It differs from a magnetron, however, in that the path of the electron beam is

determined electrostatically. The beam moves in circular paths between two concentric cylinders. The inner cylinder, which is at a higher potential than the outer one, forms part of an H.F. circuit. In this respect also it differs from a magnetron, where the outer cylinder forms part of the resonant circuit. The maximum beam current and the oscillator frequency are calculated as functions of the electron velocity, and the results are compared with experimental observations. It is amusing to notice that the electrons behave as the electric analogue of planetary satellites: the orbiting electrons are in equilibrium with the attractive force of the electrostatic field.

2530: J. J. Balder: Illuminated borders to picture screens (Light and Lighting **50**, 245-250, 1957, No. 8).

A series of tests was carried out in which a person looking at a picture screen was asked to give his opinion as to the ideal luminance and width of a uniformly illuminated border round the picture giving optimum viewing comfort. The observations were made by 20-25 persons for a number of luminance levels of both screen an surroundings. (See also Philips tech. Rev. **19**, 156-158, 1957/58 (No. 5).)

2531: P. Westerhof and J. A. Keverling Buisman: Investigations on sterols, IX. Dihydroderivatives of ergocalciferol (Rec. Trav. chim. Pays-Bas **76**, 679-688, 1957, No. 8).

Some partially hydrogenated derivatives of ergocalciferol are described; their structures are discussed, and their activities in raising the serum calcium level of rats are reported.

2532: F. J. Mulder, J. R. Roborgh, Th. J. de Man, K. J. Keuning and K. H. Hanewald: A chemical routine determination of vitamin D; correlation with the biological determination (Analysis of fat soluble vitamins II) (Rec. Trav. chim. Pays-Bas **76**, 733-746, 1957, No. 8).

A chemical routine method for the determination of vitamins D_2 and D_3 in a number of preparations is presented. The results obtained agree very well with those obtained with the rat and chick assays respectively.

2533: K. J. Keuning, G. J. van Dijk and M. J. Wiggers de Vries: Determination and adjustment of the activity of adsorbents for chromatography and of the eluting power of elution solvents by means of the "shake test" (Analysis of fat soluble vitamins III) (Rec. Trav. chim. Pays-Bas **76**, 747-756, 1957, No. 8).

Adjustment of the activity of adsorbents for chromatographic use (e.g. Al_2O_3 or $CaHPO_4$ by the controlled take-up of water) can be done in a reproducible and accurate manner with the aid of the "shake test". It consists in shaking a weighed sample (2 g) of the adsorbent with a measured volume (10 ml) of the elution solvent in which the substance to be eluted (e.g. 1000 I.U. of vitamin A) has been dissolved. The activity is suitable if about 50% is adsorbed. The following operations can be performed by means of the shake test: 1) determination of the activity of adsorbents; 2) adjustment of the activity of an adsorbent; 3) replacement of an adsorbent by another of the same activity; 4) finding the correct composition of the elution solvent; 5) replacement of one solvent by another of equal eluting power. The shake test has been developed for a 10 cm column containing 10 g of Al_2O_3; 40% adsorption is satisfactory for rapid elution. If slower elution is wanted, or if thinner columns (containing less adsorbent) are employed, the percentage adsorption must be larger than 40%, up to 60%. The shake test is a purely practical test; no theoretical significance can be attached to it.

2534: A. C. van Dorsten and J. H. Spaa: A high-output D-D neutron generator for biological research (Nucl. Instr. **1**, 259-267, 1957, No. 5).

A description is given of an apparatus capable of producing a fast neutron flux exceeding 10^{10} neutrons per second from a D-D reaction. The set consists of a pressurized cascade generator and an accelerator tube with a rotary heavy-ice target. The rectifiers used are of the selenium type.

2535: F. W. Klaarenbeek and M. H. de Lange: Enkele aspecten van de warmtetechniek van glassmeltovens (Ingenieur **69**, W137-W140, 27 Sept. 1957). (Some aspects of heating techniques in glass furnaces; in Dutch.)

A brief description is given of the melting process of glass and of the furnaces mostly used. An attempt is made to formulate the requirements of these furnaces. Some experiences from industrial practice are compared with the results of flame research at IJmuiden (Netherlands).

2536: J. Volger: Dielectric loss in insulating solids caused by impurities and colour centres (Disc. Faraday Soc. No. 23, 63-71, 1957).

Experiments at low temperatures on solids with various lattice defects have revealed the existence of dielectric relaxation phenomena due to these defects. The relaxation times are governed by activation energies far smaller than those normally found with diffusion or migration of ions. Typical measurements are given and discussed qualitatively in relation to models of some lattice imperfections including colour centres.

2537: K. Compaan and Y. Haven: Some fundamental aspects of the mechanism of diffusion in crystals (Disc. Faraday Soc. No. 23, 105-112, 1957).

Diffusion data, if available in the form $D = D_0 \exp(-E/kT)$, yield two parameters, whereas for a molecular description more quantities are needed. Some difficulties encountered with this problem are discussed. This problem has kinetic as well as equilibrium aspects. The equilibrium aspect has been successfully attacked by the theory of lattice defects: it is possible to specify the atoms that are mobile as well as the frequency with which these mobile atoms jump. Recently, a novel kinetic aspect of the mechanism has become of interest, viz., the problem of correlation between the directions of consecutive jumps of an atom. These correlations effectively alter the self-diffusion coefficient in certain mechanisms, but not the diffusion coefficient for drift or the ionic conductivity. The latter aspect has recently been developed to a promising method of disentangling diffusion mechanisms, especially in ionic crystals.

2538: E. J. W. Verwey: Onderzoekingen over oxydische ijzerverbindingen (Versl. gew. Verg. Afd. Natuurk. Kon. Ned. Akad. Wet. **66**, 106-110, 1957, No. 7). (Investigations on oxidic iron compounds; in Dutch.)

Short survey article on investigations on hexagonal ferrites carried out in recent years at the Philips Research Laboratories. These ferrites, apart from various small metal ions, always contain at least one type of metal ion of larger radius, e.g. Ba^{2+}. The crystal structures of a number of compounds of the ternary system BaO-Fe_2O_3-MeO are discussed (Me is a metal with small ionic radius). All these compounds have structures closely related to close-packed structures of the large ions (O^{2-} and Ba^{2+}). The unravelling of these structures has been accomplished by X-ray crystallographic methods. The magnetic properties of these compounds are the most important from the point of view of practical applications. (See Philips tech. Rev. **13**, 181-193, 1951/52, and **18**, 145-154, 1956/57.)

2539: L. A. Æ. Sluyterman: The amperometric titration of sulfhydryl groups with silver nitrate (Biochim. biophys. Acta **25**, 402-404, 1957, No. 2).

Three out of five sulfhydryl-containing compounds, subjected to amperometric titration with silver nitrate, appear to combine with more Ag than corresponds to their sulfhydryl content.

2540: K. Teer: Colour television transmission — practical aspects of the two-sub-carrier system (Electron. Radio Engr. **34**, 280-286 and 326-332, 1957, Nos. 8 and 9).

In recent years a colour-television transmission system using two sub-carriers for the transmission of the chrominance information has been developed. Practical results were demonstrated to Study Group XI of the C.C.I.R. in 1955 and 1956. The principles and evolution of the system have been described and discussed in previous articles. In this article the technical aspects are considered in more detail and data are given about modifications and improvements which have since been introduced.

2541: J. J. Balder: Erwünschte Leuchtdichten in Büroräumen (Lichttechnik **9**, 455-461, 1957, No. 9). (Preferred values of luminances in office rooms; in German.)

A series of tests has established what luminance values the working surface, walls and ceiling should have in office rooms in order to make the surroundings as agreeable as possible to work in. The illumination system used during the tests was ideal in the sense that light falling on the working surface comes from the right directions, and the field of vision contains no disagreeably high luminances of lamps or reflections.

2542: L. F. Defize: Fysische verschijnselen in de lasboog (De metallurgie van het lassen van staal, Lassymposium 1957, published by Ned. Ver. voor Lastechniek, The Hague, pp. 5-11). (Physical phenomena in the welding arc; in Dutch.)

The phenomena involved in gas discharges at atmospheric pressure are of a complex nature. Although the application of arc welding has grown enormously in the last twenty years, only a very limited number of papers dealing with the theory of the arc have appeared. In the arc there is the extra complication that the arc is continuously

interrupted by the droplet transfer. Most investigations on the arc discharge have therefore been made using non-consumable electrodes (carbon arc welding and gas-shielded tungsten arc welding). This paper discusses the cathodic and anodic mechanisms, the questions of the extremely high temperature of the arc and the heating effects at the electrodes. In addition to the D.C. arc, welding with A.C. is also discussed and finally the mechanism of droplet transfer is briefly considered.

2543: J. D. Fast: De rol van gassen bij het booglassen van staal (De metallurgie van het lassen van staal, Lassymposium 1957, Ned. Ver. voor Lastechniek, The Hague, pp. 12-17). (The part played by gases in the arc welding of steel; in Dutch.)

In arc welding of steel with coated electrodes, oxygen, nitrogen and hydrogen are taken up by the carbon-containing weld metal, the amounts taken up being mainly determined by the composition of the coating. In general no equilibria are established and the amounts of the gases taken up cannot therefore be calculated, but nevertheless the laws of chemical equilibrium do enable us to draw a few important conclusions of a qualitative nature. Two of the most important conclusions are: 1) the more stable the oxides of the coating, the less oxygen will be absorbed by the metal; 2) if the water content of the coating remains constant, the amount of hydrogen taken up by the metal will be greater the smaller the amount of oxygen taken up. When the metal solidifies, part of the C, O, N and H present is given off in the form of CO, CO_2, H_2O, H_2 and N_2. Under adverse conditions the evolution of these gases may lead to porosity of the weld. The role played by sulphur in this connection is discussed. The deleterious effects of oxygen, nitrogen and hydrogen left in the metal after solidification, especially aging effects, are also discussed. On the basis of dislocation theory, tentative conclusions can be arrived at regarding the influence of interstitial elements on the brittle fracture of steel and on the different sensitivities of ferritic and austenitic steels to these impurities.

H 1: F. Karstensen: Preferential diffusion of Sb along small-angle boundaries in Ge and the dependence of this effect on the direction of the dislocation lines in the boundary (J. Electronics and Control **3**, 305-307, 1957, No. 3).

Note reporting experiments which show that the rate of diffusion of antimony along the dislocation lines forming a small-angle boundary in germanium exceeds that of self-diffusion in germanium. No such effect is observed in the crystal boundary perpendicular to the dislocation lines.

NOW AVAILABLE

J. D. Fast, H. G. van Bueren and J. Philibert (editors): La diffusion dans les métaux (Philips Technical Library, 1957; pp. 115, 84 figures).

This book (published only in French) forms a record of the papers presented at a colloquium held at Eindhoven in September 1956 on the subject of diffusion in metals. The first paper gives an introductory survey and, of the following papers, two are concerned with the effect of structural defects on the diffusion, four on the Kirkendall effect, two on the diffusion of interstitial atoms and one on the effect of elastic stress on diffusion. The English titles of the papers are as follows: Introduction to the study of diffusion, by A. D. le Claire; The diffusion of radio-tracers in solids, by K. Compaan and Y. Haven; Intergranular diffusion and its relation to grain boundary structure, by P. Lacombe; Intergranular self-diffusion of α-iron, by C. Leymonie and P. Lacombe; New observations on the Kirkendall effect and on electrolytic transport in solid alloys, by Th. Heumann; An electron-beam micro analyser and its use in the study of intermetallic diffusion, by J. Philibert; Study of the diffusion uranium-zirconium in the γ phase, by Y. Adda and J. Philibert; The Kirkendall effect and diffusion in the system gold-platinum, by A. Bolk and T. J. Tiedema; Diffusion of interstitial atoms, by J. L. Meijering; Study of the non-steady permeation of helium through silicon and germanium by means of a mass spectrometer, by A. van Wieringen; Effect of elastic deformation on the mobility of vacancies in copper, by C. W. Berghout.

VOLUME 19, 1957/58, No. 12 pp. 333-364 Published 22nd August 1958

Philips Technical Review

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

THE DOUBLE-FLUX "TL" LAMP, A FLUORESCENT LAMP OF HIGH OUTPUT PER UNIT LENGTH

by H. J. J. van BOORT and D. KOLKMAN.

621.327.534.15

A minor modification in the design of Philips tubular fluorescent lamps has made it possible to attain, with a satisfactory efficiency, about double the luminous flux normally obtained in fluorescent lamps of the same length. For a variety of purposes, such as the lighting of factory workshops, sports halls, etc., these new lamps will make possible more economical installations.

Although the fields of application of incandescent lamps and tubular fluorescent lamps ("TL" lamps) are now fairly clearly defined — they enjoy a sort of peaceful co-existence — we are reminded from time to time of some valuable properties of the incandescent lamp that are missing in the fluorescent lamp. One of these properties is the concentration of a large output in a relatively small and easily manipulated lamp. Incandescent lamps up to 1500 W are still classed as "ordinary" lamps, i.e. as lamps requiring no special methods and procedures in manufacture or in handling (for connection, maintenance and replacement). Far larger incandescent lamps, up to 10 kW or even more, are regularly manufactured and even these can be reasonably manipulated. The world success of the fluorescent lamp, on the other hand, has been achieved largely with a single version, viz. that with an output of approx. 40 W and a length of about 1.20 m. Until recently it seemed that a substantially larger power could not be obtained in a lamp of this length without largely sacrificing the advantage of its high luminous efficiency.

This restriction in power was an unsatisfactory state of affairs. For high light level industrial and street-lighting installations, for which fluorescent light is so eminently suitable as regards colour and efficiency, it is most desirable that the luminous flux of each lamp be as high as possible: fewer lamps are then required, costs of fixtures and ballasts are thereby lowered and maintenance is cheaper. The rather trivial measure of making the lamps longer whilst retaining about the same power per unit length, however, is not a real solution at all. The fixtures may be fewer in number, but will also be more expensive, and the same applies to the ballasts, owing to the ignition voltage increasing with the length of the lamp. Apart from that comes the fact that replacing very long lamps, for instance 125 W lamps with a length of 2.40 m, in an installation for industrial or street lighting is no easy matter.

Lamp manufacturers have now succeeded, in various ways, in developing fluorescent lamps with a larger output, whilst retaining a reasonable length and a high luminous efficiency. In order to appreciate the modifications introduced for this purpose, let us first examine the factors determining the efficiency of the fluorescent lamp.

Efficiency and dimensions of tubular fluorescent lamps

The basic process in lamps of this type is the emission of ultraviolet radiation quanta, mainly at a wavelength of 2537 Å, by mercury atoms in an excited state, and the absorption of these quanta by the fluorescent powder deposited on the tube wall. The larger the number of mercury atoms present, i.e. the higher the mercury-vapour pressure in the tube and the larger the current density of the electric discharge, the more mercury atoms will be excited by the electrons accelerated in the electric field (the primary process). At the same time, however, two secondary phenomena which impair efficiency become increasingly manifest: these are self-absorption, i.e. the absorption of radiation quanta by unexcited mercury atoms, and the raising of excited mercury

atoms to higher energy levels. In either case, it is true, a 2537 Å radiation quantum may be re-emitted, but the competing processes, namely the emission of quanta of larger or smaller wavelength, which are unsuitable for producing fluorescence, and the dissipation of excitation energy in the form of heat, are favoured by these intermediate excitation stages [1]). Particularly adverse is the effect of a rise in mercury-vapour pressure. On the other hand, some mercury pressure is obviously essential to the primary process. An optimum light efficiency is found at a specific mercury-vapour pressure, viz. at approx. 5×10^{-3} mm Hg. This corresponds to the saturation vapour pressure of mercury at a temperature of about 40 °C.

"TL" lamps (being provided with a small surplus of liquid mercury) are accordingly always made in such a way that the tube wall will assume a temperature of about 40 °C when the lamp is burning under ordinary conditions. It is the wall of the tube, being the coldest part of the lamp, which determines the equilibrium pressure of the mercury vapour. In ordinary "TL" lamps the tube wall assumes this temperature along nearly the whole discharge column; near the electrode it is generally warmer.

Once the wall temperature of the discharge column is specified, a certain relationship will hold between the power to be developed in the discharge column and the dimensions of the tubular lamp: for a given ambient temperature, generally taken as 25 °C [2]), and for a given wall temperature, the amount of heat q dissipated per unit time from each unit area of the wall by radiation and convection, may be considered as a fixed value. In actual fact, this amount q depends to some degree upon the diameter d of the tube (q decreases as d is made larger), but this may be disregarded to a first approximation. Consequently, if the discharge column is to handle a power P, a wall area $A = P/q$ is required.

Let us now assume that we wish to double the power developed in the column. This means that also the wall area has to be approximately doubled. One obvious way to do so is as already mentioned, viz. by making the column twice as long with the same diameter. This has more or less the same effect as if two of the original lamps were placed end to end. We have already mentioned that this is no adequate solution: the lamp becomes undesirably long and also the voltage V_z across the column becomes far too high, viz. twice the original value. An effective lamp design is one for which the burning voltage (i.e. arc voltage V_z + the voltage loss at the electrodes, the latter being about 15 V for "TL" lamps) is approximately half the r.m.s. value of the mains voltage (or, more accurately: of the open voltage of the ballast). A higher burning voltage is not favoured, because then the lamp would be unduly affected by mains-voltage fluctuations. This could be compensated by stepping-up the mains voltage, but that would again lead to a more expensive ballast.

Another method of obtaining double the wall area would be to double the diameter whilst retaining the original length of the column. This would be a good solution, where easy handling of the lamp is concerned, but the overall efficiency would be very poor. With such a diameter the arc voltages become very low, whilst the current becomes very large (the arc-discharge gradient decreases with increasing diameter and with increasing current, see below). For a very low arc voltage the contribution of the constant voltage loss of 15 V at the electrodes would impair the total efficiency to such an extent that it would more than offset the advantage of having maintained the correct wall temperature and accordingly the optimum mercury-vapour pressure. The large current and low voltage, moreover, would again require a relatively expensive ballast.

An obvious compromise between these two extreme possibilities (constant diameter d or constant length l) would be to double the wall area $A = \pi dl$ whilst retaining the burning voltage. Both length and diameter must then be greater than those of the original lamp. The required values can be approximately found with the aid of a graph giving the experimentally determined gradient of the low-pressure mercury discharge as a function of the diameter, for various values of the current; see *fig. 1*. The procedure is explained in the caption to the graph. The result is a roughly geometrically-similar enlargement of the original lamp: both length and diameter have to be enlarged by a factor of about $\sqrt{2}$ to double the power; see *fig. 2*.

This may be illustrated by the following numerical example. The well-known "TL" 65 W lamp, which has been manufactured for some years now, has a burning voltage of 110 V and hence a voltage $V_z = 110 - 15 = 95$ V across the column; the power developed in the column is approximately 55 W. The internal diameter of the lamp is 35 mm and its length about 1.50 m; for the electrode spaces and the transition zones between column and electrode about 10 cm should be allowed, so that

[1]) See e.g. G. Heller, Comparison between discharge phenomena in sodium and mercury vapour lamps, Philips tech. Rev. 1, 2-5 and 70-75, 1936.

[2]) In accordance with the international agreement to measure the luminous flux and hence also the luminous efficiency of fluorescent lamps always at an ambient temperature of 25 °C.

the column has a length of approx. 1.40 m. Doubling the power developed in the column to 110 W, according to fig. 1, would lead to a lamp having a column length of approx. 2.20 m and an internal diameter of 48 mm.

The foregoing evidently does not reveal what the consequences are if we depart from the prescribed relation between l and d, i.e. to what extent the luminous efficiency of the lamp would be impaired if the wall temperature were allowed to assume a

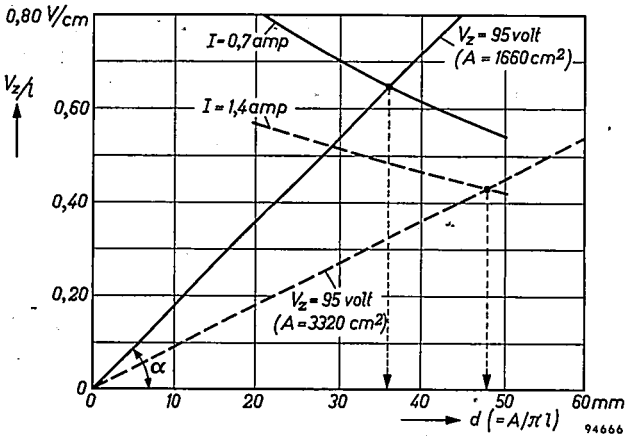


Fig. 1. The curved lines show how the gradient V_z/l in the column of a low-pressure mercury discharge varies as a function of the internal diameter d, for various discharge currents I (for an argon pressure of 1.5 mm Hg).

Given a certain wall temperature, then for a prescribed power P developed in the column, the wall area A may be considered roughly as fixed; hence the abscissa scale for d $(= A/\pi l)$ at the same time represents a reciprocal scale of the column length l. On any straight line through the origin, therefore, the voltage V_z has a constant value, and the required value of d (and hence of l) for given values of P and V_z is given by the intersection of the curve corresponding to the current I with the straight line appropriate to V_z.

For a larger current, corresponding to a larger power P (and therefore larger A), the V_z/l curve lies lower (dotted curve); also the slope of the straight line for the same value of V_z is less. (The slope of this straight line is $\tan \alpha = \pi V_z/A$, and is therefore, at constant V_z, smaller for the larger area A corresponding to the higher value of P.) The new point of intersection occurs at a larger diameter and length, these being roughly proportional to the square root of the power.

value different from the optimum 40 °C. A quantitative evaluation of this effect would exceed the scope of this article, the more so since we have to consider that this efficiency does not depend on the mercury pressure alone. The influence of the current density we have mentioned earlier; additional factors are the pressure of the rare gas present to aid in the ignition of the lamp, the length of the free path of the radiation quanta from the interior of the column to the wall, the recombination of ions and electrons at the wall, etc. The simplest way of demonstrating the importance of the correct wall temperature, that is to say, of the correct specific loading, is to carry out an experiment. For this purpose we have measured the luminous efficiency of a given ordinary "TL" lamp as a function of the load; see *fig. 3*. We

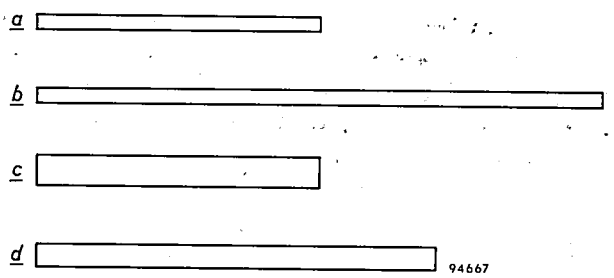


Fig. 2. Given a tubular fluorescent lamp of a certain power, a lamp with double the power may be realized in various ways:
a) Original lamp.
b) Lamp with the same diameter and double the length; the arc voltage is doubled.
c) Lamp having the same length as (*a*) but with double the diameter; the arc voltage becomes very low.
d) Lamp having the same arc voltage as (*a*); both length and diameter have been increased; the proportions of the lamp are similar to (*a*).

then find that by increasing the power to double the rated value, the luminous flux is raised by only 40%. The luminous efficiency of the lamp has therefore dropped by 30%. A noticeable reduction in luminous efficiency (viz. by 5%) is even apparent when the rated power is exceeded by as little as 16%.

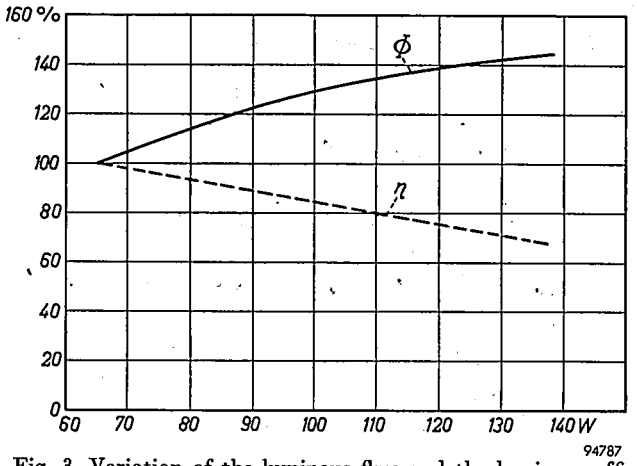


Fig. 3. Variation of the luminous flux and the luminous efficiency (both plotted as percentages of the rated values) of a given "TL" 65 W lamp for various values of the power developed in it. The lamp was not of course operated with its ordinary ballast: it was fed from a special supply. (The lamp differed from the ordinary "TL" 65 W lamps only in having somewhat heavier electrodes, fitted for the purpose of this experiment to withstand the larger current.)

Fluorescent lamps for increased power per unit length

The foregoing considerations led to the conclusion that for a larger power a longer fluorescent lamp would be necessary. On further consideration, however, we find that this is due to the implicit assumption that the temperature determining the mercury-vapour pressure in the tube (the temperature of the coldest spot of the tube) should be the same as that determining the amount of heat carried off per cm²

of the wall (some form of average wall temperature). This provides the clue showing the direction one must take in order to develop a larger power in a fluorescent lamp without appreciably increasing its length: the design of the lamp should be changed such that, *with a higher average wall temperature, there remains at least one spot where the temperature of the wall is about 40 °C.*

We shall now discuss three designs of fluorescent lamps in which this principle has been employed: the "Power Groove" lamps of General Electric, U.S.A., the "V.H.O." (very high output) lamps of Sylvania, and the "TL" double-flux lamps developed by Philips.

The design of the "Power Groove" lamps [3]) is based on a tube with a diameter of 53 mm, the cross-section of which has been deformed into the kidney-shape shown in *fig. 4*. Since a groove running over the entire length of the lamp would reduce the

The "V.H.O." lamps [4]) have a normal circular cross-section with a diameter of 38 mm, but for their given total length of 1.20 m the distance between the electrodes is smaller than normal. Thus,

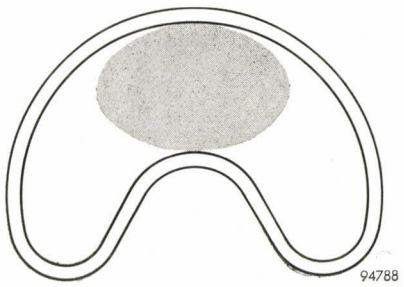


Fig. 4. Cross-section of the "Power-Groove" lamp (General Electric [3])). The discharge is more or less concentrated in the shaded part; in the lower points of the cross-section the wall remains at a lower temperature.

mechanical strength, thus increasing the risk of implosion, the lamp has been allowed to retain some lengths of circular cross-section at the ends and at a few regular intervals in between. The discharge mainly concentrates in the shaded part of the cross-section. The lower parts of the cross section (the lamp is mounted in the position shown in fig. 4) remain relatively cool. Having a length of 1.20 m, such a lamp can be loaded with 107 W (current 1.5 A at a voltage of 84 V) and then has a luminous efficiency of 56 lm/W, measured after 100 burning hours.

behind each electrode remains a space with a length of about 6 cm in which no discharge takes place (*fig. 5*). This space is moreover screened off from the radiant heat of the electrode by a metal plate, and is thus more or less separated from the discharge section. This makes it possible to keep the wall temperature at the end of the tube at about 40 °C, with a load of 100 W. Neon is used as the ignition gas instead of the customary argon.

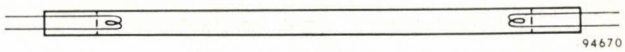


Fig. 5. Sketch of the "Very High Output" lamps (Sylvania [4])). Here the sections situated behind the electrodes, which are longer than normal and screened off from the heat radiated by the electrodes, remain at a temperature of 40 °C.

Current and burning voltage of this lamp amount to 1.2 A and 99 V.

The Philips-designed new type of "TL" lamp for large output per unit length, the "TL" double-flux lamp, is shown in *figs. 6* and *7*. This lamp also has the ordinary circular cross-section, with the exception, however, of a small protrusion made at half its length, such that the temperature at the bottom of the protrusion remains at about 40 °C, whilst all other parts of the wall become appreciably warmer. One version made on this principle has a length of 1.50 m, i.e. the same length as the "TL" 65 W lamp, and a diameter of 35 mm. With a protrusion of diameter 10 mm and depth 10 mm, the tube can be loaded with 125 W, i.e. the same load per unit length as that of the two above-mentioned American types. The current is then 1.5 A, the burning voltage 97 V, whilst the luminous efficiency after 100 hours amounts to 57 lm/W. The lamp therefore develops per unit length a luminous flux that is 1.7× as large as that of the "TL" 65 W lamp, which has an efficiency of 64 lm/W, and 2.2× as large as that of the "TL" 40 W lamp, whose efficiency amounts to 66 lm/W. The tube wall (apart from the protrusion and the regions near the electrodes) assumes a temperature of 57 °C at an ambient temperature of 25 °C. The fact that the luminous efficiency is somewhat lower than that of the "TL" 65 W lamp, although the optimum mercury-vapour pressure is maintained, is mainly attributable to the higher current density, which promotes the excitation of mercury atoms to higher energy levels.

The above design with the larger power per unit length may be applied for making lamps of various

[3]) J. O. Aicher and E. Lemmers, Design and characteristics of fluorescent lamps having a non-circular cross section, Illum. Engng. **52**, 579-584, 1957 (No. 11). G. R. Baumgartner, R. T. Dorsey and E. A. Linsday, Application of non-circular cross-section fluorescent lamps, Illum. Engng. **52**, 587-596, 1957 (No. 11).

[4]) J. F. Waymouth, W. Calvin Gungle, Ch. W. Jerome and F. Bitter, Very-high-output fluorescent lamps, Sylvania Technologist **9**, 102-110, 1956. J. F. Waymouth, F. Bitter and E. F. Lowry, Factors to be considered in the design of high-output fluorescent lamps, Illum. Engng. **52**, 257-261, 1957 (No. 5).

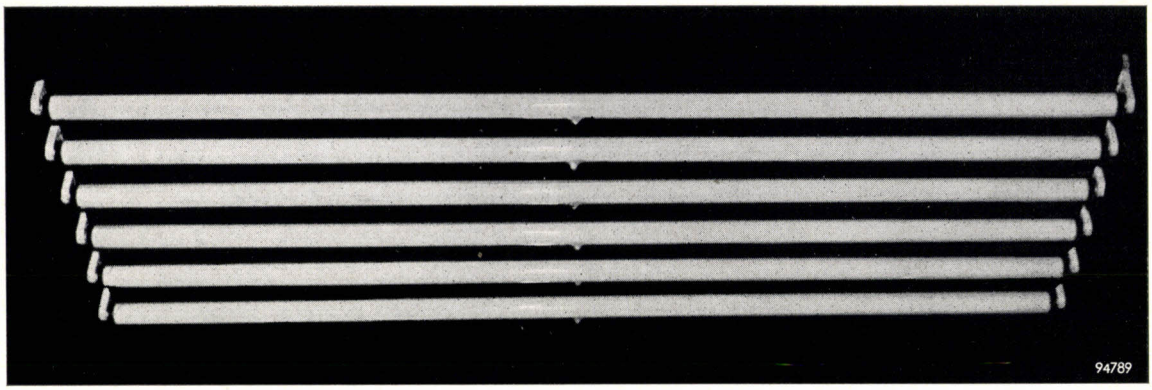


Fig. 6. A row of "TL" double-flux lamps.

power ratings (and hence lengths). The version described here, with a length of 1.50 m, has been preferred in order to attain a relatively high burning voltage. The use of neon instead of argon as ignition gas would likewise result in a higher burning voltage, but was found to be less desirable as it induced a more rapid rate of vaporization of the electrodes and hence a shortened working life. With a view to this vaporization we have also preferred a higher argon pressure, viz. $\approx$ 1.5 mm Hg, rather than the pressure of about 1 mm Hg prescribed for optimum efficiency.



Fig. 7. Sketch of the "TL" 125 W double-flux lamp developed by Philips; the length is 1.50 m (the same length as the "TL" 65 W). There is a protrusion of depth and diameter about 10 mm on the tube half-way along its length. The temperature in the protrusion does not exceed 40 °C, so that the mercury-vapour pressure in the tube (after all the surplus mercury has condensed here) eventually assumes the optimum value.

The situation of the "cold spot" half-way along the length of the tube has the advantage of being least susceptible to the high temperature around the electrodes, so that a protrusion of such small dimensions is adequately effective. This design has the added advantage that the lamp emits light without brightness variations along its entire length. The protrusion can be made without difficulty *after* the tube has been provided with its fluorescent layer, thus avoiding any problems in depositing the latter.

After the lamp has been filled with inert gas and mercury and sealed off, the liquid mercury is distributed at random inside the lamp. Once the lamp has been delivered to site, mounted in its position and ignited for the first time, it will burn for some time at too low a voltage, because the mercury-vapour pressure is too high (there are drops of liquid mercury on the hot walls). It will take a number of burning hours before all the surplus mercury is distilled from the hot regions and condensed in the protrusion, giving the correct equilibrium mercury-vapour pressure. Only then will the burning voltage attain the required value and the lamp yield its maximum luminous flux. The period necessary for this stabilization differs from lamp to lamp, since the initial distribution of the mercury may vary a great deal.

After this stabilization period the liquid mercury will remain in the protrusion, so that almost immediately after the lamp is switched on again the maximum luminous flux is obtained. It is obvious that the lamp should only be mounted horizontally, with the protrusion in the lowermost position.

Summary. Tubular fluorescent lamps ("TL" lamps) have the highest luminous efficiency at a mercury-vapour pressure of about 5×10^{-3} mm Hg. With ordinary lamps this pressure is obtained because the temperature of the tube wall assumes a value of about 40 °C. It can be demonstrated that under this condition and for a given burning voltage the power can only be increased by making the lamp longer. However, fluorescent lamps have now been developed in which the larger part of the wall can assume a higher temperature, whilst a minor part of the area remains at 40 °C; in this case the optimum mercury-vapour pressure is retained. The "TL" double-flux lamp designed by Philips on this principle, and characterized by a small protrusion, about 10 mm in diameter and 10 mm deep, midway along the tube, has a power of 125 W at a length of 1.50 m, i.e. the same length as the ordinary "TL" 65 W lamp. The luminous efficiency of the new lamp is 57 lm/W compared with 64 lm/W for the ordinary "TL" 65 W lamp, both measured after 100 burning hours.

Erratum. *A simple apparatus for contact microradiography between 1.5 and 5 kV*, by B. Combée and A. Recourt. (Philips tech. Rev. **19**, 221-233, 1957/58, No. 7-8).
On p. 227, column (ii) of the above-mentioned article, the second sentence of the last paragraph should read:
"The experimental emulsion differs from the standard "Maximum Resolution" emulsion (type V 6005) in that it is only **4** microns thick instead of **5**, and has **5** times more grains per unit volume; the size of the AgBr grains is the same".

THE PROJECTION OF COLOUR-TELEVISION PICTURES

by T. POORTER and F. W. de VRIJER. 621.397.2:621.397.62:535.881

A projection system for monochrome television was developed in the Philips Eindhoven Laboratory over 10 years ago. Further work has now resulted in the development of apparatus for the projection of coloured television pictures on to a screen. Projectors have been built giving pictures 22 cm × 29 cm and 35 cm × 46 cm at a highlight luminance of 200 cd/m², and 2.25 m × 3 m at a highlight luminance of 20 cd/m². The latter value is about the same as that on normal cinema screens. The projectors were demonstrated before the colour-television commission of the C.C.I.R. in 1955 and 1956.

Almost all colours occurring in nature can be simulated by adding together three kinds of light, each of a suitable primary colour: red, green and blue. Colorimetric theory, as it concerns the addition of colour, was dealt with in an earlier article [1]) which will be referred to as I. It is possible to obtain a television picture in natural colour in the manner stated by building up each element of the picture out of appropriate quantities of light of the three primary colours. A practical method of doing so is to project a red, a green and a blue image simultaneously on to a common screen, in such a way that the three images coincide exactly. The luminance of each element of the resultant image is thus the sum of the luminances of the three corresponding elements of the primary-colour images, the colour of the resultant element being determined by the proportions of these luminances.

One obvious way of building such an apparatus would be to combine a projection type cathode-ray tube giving coloured images — working on the same principle as direct-viewing tubes for colour television — with a suitable optical system, but no projection tube for coloured images has so far been produced. However, projection tubes luminescing in each of the three primary colours do exist. A set of three projectors, one for each of the three primary colours is therefore taken as the basis of the system: the three projectors are equipped with projection tubes giving red, green and blue light.

It is of course necessary that corresponding elements of the primary-colour images should coincide on the projection screen. The correct superposition of the three images is therefore a matter requiring particular attention. One of the conditions for proper superposition is that the three projectors should be identical in design and that any differences arising in their construction should be kept as small as possible.

The primary-colour projectors

The primary-colour projectors can be designed on exactly the same lines as those for monochrome television. A projector of this kind, embodying the Schmidt optical system, was described in this journal [2]) and elsewhere [3]) in 1948. The Schmidt system combines three valuable properties, namely low optical distortion, low transmission losses and large aperture. In its simplest form the system comprises the arrangement shown diagrammatically in *fig. 1a*. However, as stated in the article referred to in footnote [2]), the alternative arrangement shown in fig. 1*b* has been used in home projection receivers. Here the light beam is "folded" by means of a plane mirror M_2, which is placed at

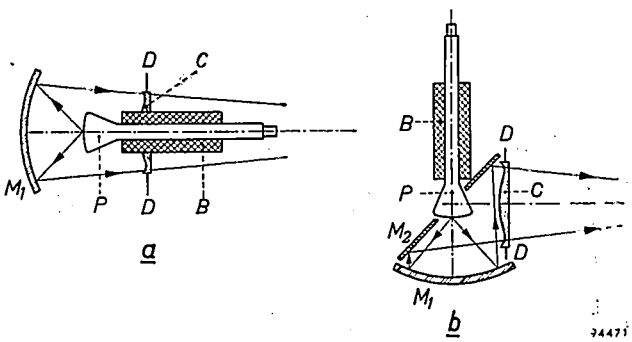


Fig. 1. Schmidt optical systems for the projection of television images; *a*) "in-line" Schmidt system, *b*) "folded" Schmidt system. *P* projection tube with focussing coil and scanning coils *B*. M_1 spherical concave mirror. *C* corrector plate mounted in diaphragm *D*. The centre of curvature of M_1 lies in the plane of *D* in the case of (*a*) and in the image of *D* given by plane mirror M_2 in the case of (*b*).

[1]) F. W. de Vrijer, Fundamentals of colour television, Philips tech. Rev. **19**, 86-97, 1957/58 (No. 3).

[2]) P. M. van Alphen and H. Rinia, Projection-television receiver, I. The optical system for the projection, Philips tech. Rev. **10**, 69-78, 1948/49.

[3]) H. Rinia, J. de Gier and P. M. van Alphen, Home projection television, I. Cathode-ray tube and optical system, Proc. Inst. Rad. Engrs. **36**, 395-400, 1948.

an angle of 45° to the principal axis. This arrangement allows the set to be constructed more compactly and to be enclosed in a dust-proof box. When space is availabe, however, there is no reason why the "in-line" optical system shown in fig. 1*a* should not be used [4]).

Superposition of the three primary-colour images

There are several possible ways of arranging for the primary-colour images to coincide on the projection screen. A general distinction may be made between arrangements whereby the three projectors, from an optical point of view, are effectively coincident (all done by mirrors), and all other arrangements.

Three projectors effectively at the same point

In theory superposition errors are absent if one can somehow place the projectors such that the three images are projected, as it were, from one point. *Fig. 2* shows how this may be done with the aid of two dichroic mirrors as discussed recently in this journal [5]). Mirror *r* reflects red light and passes green and blue; mirror *b* reflects blue and passes red and green. Green light has to go through both mirrors, the proportion transmitted being between 70% and 80%.

The crossed mirrors may be mounted with the line of intersection either horizontal or vertical. If the line of intersection is horizontal the three projectors will have to be in a vertical plane, and

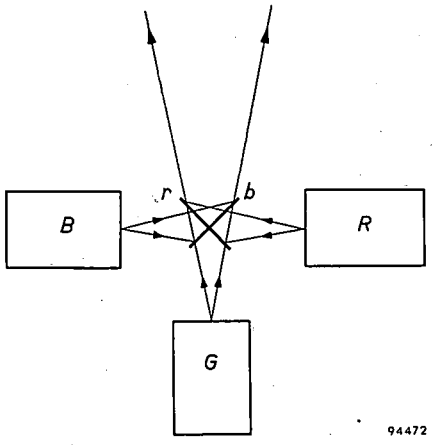


Fig. 2. Superposition of three primary-colour images by means of crossed dichroic mirrors (*r* and *b*). Mirror *r* reflects the red light from *R* and passes the green from projector *G* and the blue from projector *B*. Mirror *b* reflects blue light and passes red and green.

[4]) J. Haantjes and C. J. van Loon, A large-screen television projector, Philips tech. Rev. **15**, 27-34, 1953/54.
[5]) P. M. van Alphen, Philips tech. Rev. **19**, 59-67, 1957/58 (No. 2).

this has the advantage that colour variations over the area of the projected picture are reduced to a minimum. The variations arise because the light beams do not all strike the mirrors at the same average angle; with the line of intersection horizontal,

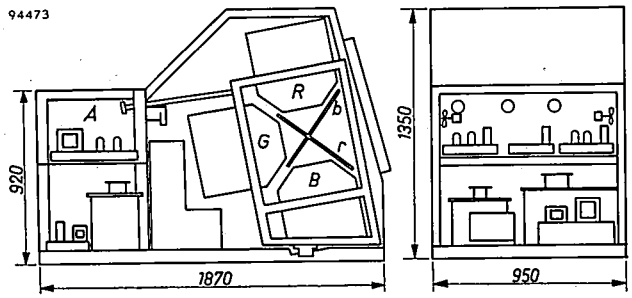


Fig. 3. Sketch of the first large-screen colour-television projector constructed at Eindhoven (see also fig. 4); it gives a picture measuring 2.25 m × 3 m (≈ 7½′ × 10′). *r* and *b* are crossed dichroic mirrors with their line of intersection horizontal. This means that *R*, *G* and *B*, the primary-colour projectors, are mounted in a vertical plane; despite this it was possible to limit the overall height of the apparatus to 1.35 m. Compartment *A* contains the supply unit. The dimensions shown are in mm.

the angle over which variations occur is subtended by the lesser dimension of the picture, namely its height. However, if the apparatus is for projecting large pictures, care must be taken not to build it too tall, otherwise it will interfere with part of the audience's view of the screen. This problem arises because in order to keep down the dimensions of the main optical system, the projectors have to be placed as close to the sreen as possible, i.e. in front of the viewers. It has proved possible to keep the height of the projection apparatus down to 1.35 m as shown in *fig. 3*, which gives the dimensions of the first of the large-screen colour television projectors built in this laboratory (photograph in *fig. 4*).

If the mirrors are mounted with their line of intersection vertical and the primary-colour projectors are accordingly placed in a horizontal plane, it is easier to keep down the height of the apparatus but there will be greater colour variations in the projected picture.

Much depends on whether the crossed mirrors are combined with "in-line" or "folded" Schmidt optical systems. In the first kind of Schmidt system the projection tube, surrounded by the focussing and scanning coils, protrudes through an opening in the corrector plate (fig. 1*a*). In order that the tube and coils should block the path of the reflected light as little as possible, they have to be located outside the crossed mirrors. These latter have therefore to be set up some distance away from the corrector plates (in fig. 1*a* they would be mounted to the right of the foot of the tube), in a place where the re-

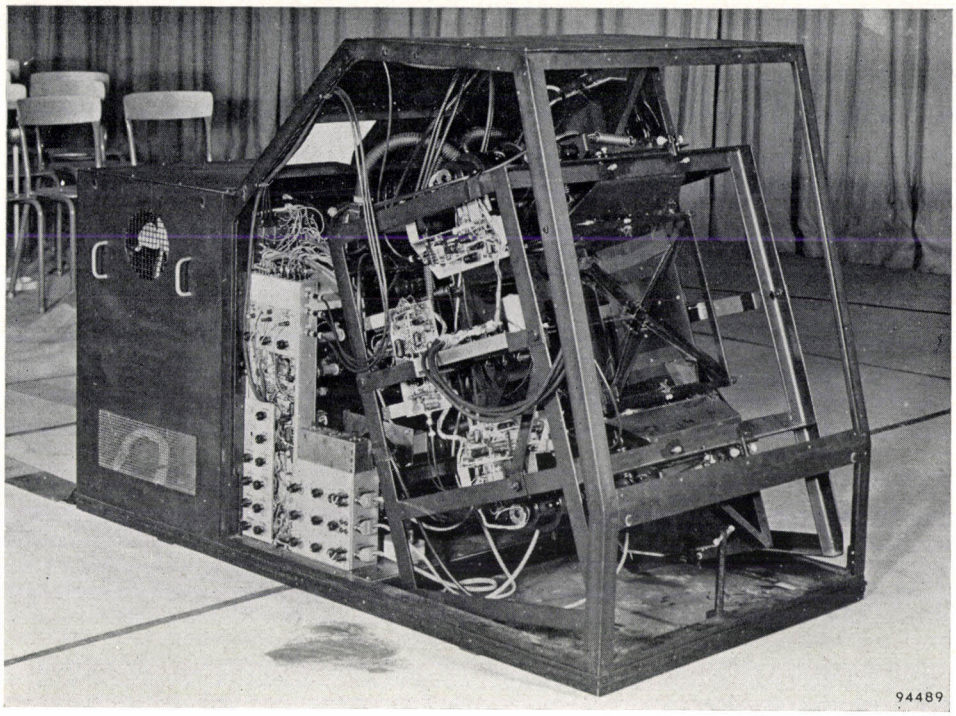


Fig. 4. Photograph of the colour-television projector sketched in fig. 3. The size of the projected picture is 2.25 m × 3 m (≈ 7½′ × 10′) and its maximum luminance is about 10 cd/m². Electrical adjustments are carried out from a control desk (not shown in the photograph).

flected beams have a fairly large diameter. In consequence, the mirrors have to be bigger than they are in "folded" systems (fig. 1*b*), where the crossed mirrors can be set up close to the corrector lenses. The "in-line" optical system also necessitates special arrangements for inserting and withdrawing the CRT; for example, it might be necessary to design the concave mirror in such a way as to allow of dismantling.

Projectors using horizontally intersecting dichroic mirrors have been built, and have given good results. In general, however, designs embodying dichroic mirrors have two disadvantages. The first is a loss of brightness and contrast, due to the large number of reflecting and absorbing surfaces in the path of the beams, especially if these are "folded". The second disadvantage is a loss of sharpness due to the difficulty in achieving a completely stress-free mounting of the mirrors. In consequence, ways have been sought of superimposing the three primary colour images without the aid of dichroic mirrors. This line of enquiry has led to arrangements whereby the three primary-colour images are *not* projected effectively from the same point. This necessarily results in certain superposition errors and, in general, some compensation of these must be sought, using special corrective measures.

Projectors not effectively at the same point

a) The projectors are placed at the apexes of an equilateral triangle. Of all the different arrangements that are possible, this involves the smallest errors of superposition; the triangular arrangement will therefore be attractive where it is not desired to undertake any corrective measures. However, such measures can only be dispensed with if the optical magnification is large — the errors diminish in relation to the picture size as the magnification increases. Correction of the errors, necessary where magnification is not so large, is more complicated in this arrangement than in those now to be discussed.

b) The projectors are arranged side by side. We may distinguish between two cases: (1) where the three projectors are parallel to one another, their axes being perpendicular to the projection screen, and (2) where the outside projectors are inclined inwards slightly.

The following observations apply equally well, *mutatis mutandis*, to arrangements whereby the three projectors are mounted one above the other. Such arrangements are less desirable, however, because they occupy a greater height.

b1) The three projectors are side by side with their axes parallel. Each throws a rectangular image on to the screen and, in theory, the projected images

can be made to coincide by displacing the luminescent images on the two outside tubes electrically. If the luminescent images are to be displaced from the central position on the tube screen it is necessary that they should not occupy the whole of its area. Thus only part of the tube screen is utilized, with the consequence that stronger magnification is necessary and that the highlight luminance on the projection screen will be lower than the maximum obtainable. In addition, vignetting will occur, owing to the fact that the central ray passes obliquely through the optical system. It is a condition for success that the deflection coils should be so designed that the images on the projection screen are completely free from distortion. This ideal cannot be approached closely enough in practice, and so the method can only be used where the magnification is very large and where, in consequence, the outside CRT screen images need only be displaced slightly.

b2) The outside projectors are pointed inwards (*fig. 5a*) in such a manner that their optical axes intersect in the centre of the projection screen. The deflection coils are adjusted such that the vertical centre lines of the separate images coincide (along *MM'* in fig. 5*b*). The result is that, for the images from the outer projectors the magnifications are not equal in the left and right halves of the picture, giving rise to trapezium shaped images. By introducing an electrical correction into both the line and frame scanning of the outside tubes (how this is done will be discussed below) the images from these two projectors can be made to coincide with the rectangular image projected by the centre projector.

Prior to correction, with the three projectors positioned in the best possible manner, the superposition errors arising in the corners of our projected pictures were as follows. In a picture measuring 2.25 m × 3 m the error was six times the inter-line separation; in a picture measuring 22 cm × 29 cm the error was thirty times the inter-line separation. Both cases refer to a raster of a 625 lines.

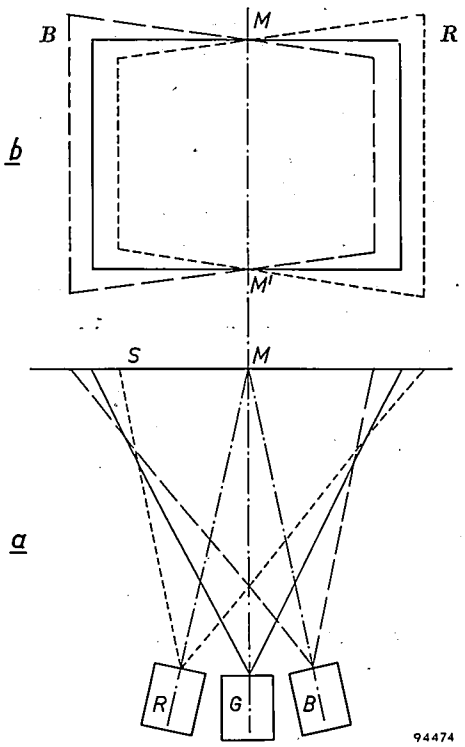


Fig. 5. *a*) Arrangement whereby the three primary-colour projectors *R*, *G* and *B* are placed side by side, with the two outer ones inclined inwards at an angle such that the vertical lines through the centres of the three images (*MM'* in diagram *b*) coincide.
b) In consequence, the red and blue images suffer from trapezium distortion.

Since the axes of projectors *R* and *B* are not now perpendicular to the projection screen (fig. 5*a*) there ceases to be a linear relationship between the horizontal movements of the red and blue spots on the screens of the outside tubes and the movements of their respective projected images. In order to give the spots on the projection screen a constant horizontal velocity, the line scanning current must be given the form indicated in *fig. 6a*. This means adding correction currents to the normal line scanning currents of the outside projection tubes. The two correction currents are, of course, equal and opposite; hence only one current has to be generated and to be passed in the appropriate

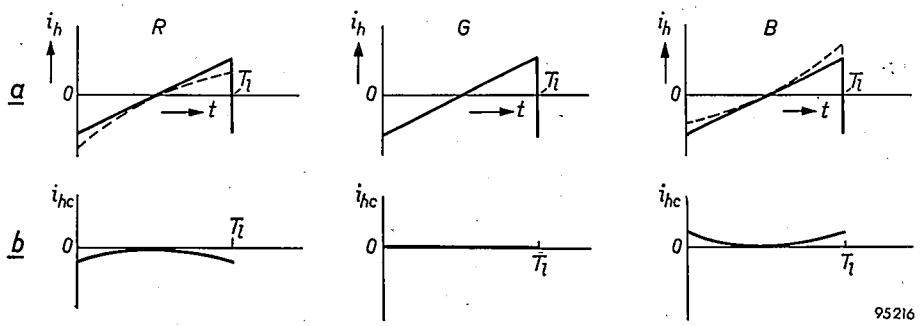


Fig. 6. *a*) Waveform required (dashed curves) for the line scanning current (i_h) in primary-colour tubes *R*, *G* and *B* (mounted as in fig. 5*a*) in order that linearity be restored to the movement of the three spots across the projection screen. T_l is the line period.
b) To obtain the above waveforms, the linear sawtooth currents in *R* and *B* (full lines in *a*) must have correction currents i_{hc} added to them; the two correction currents are equal and opposite.

direction through the scanning coil of each of the outside tubes. Prior to this correction, the three images on the projection screen have the outlines shown in *fig. 7a* after applying the above correction, they take on the shapes shown in fig. 7*b*.

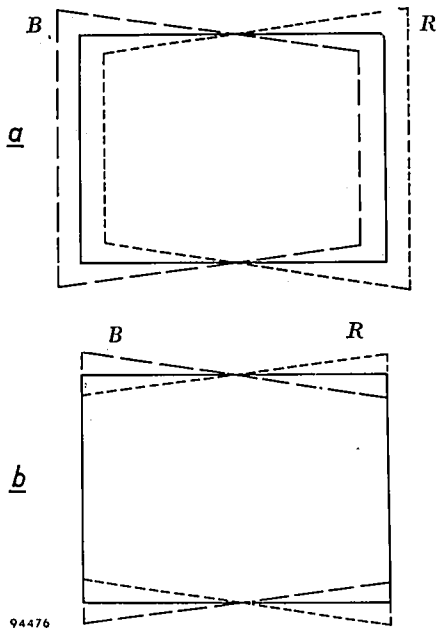


Fig. 7. *a*) Outlines of the three primary-colour images on the projection screen prior to correction (cf. fig. 5*b*).
b) Application of the correction illustrated in fig. 6 brings the vertical edges of the three images into coincidence. To get complete coincidence of the images, correction of the vertical (frame) scanning is further necessary — see fig. 8.

The red and blue images on the projection screen are still of trapezium shape; to make them coincide with the rectangular green image, the frame scanning must also be subjected to a correction. From a glance at the outline of, say, the blue image in fig. 7*b*, it will be clear that the vertical scan has to be shortened on the left and lengthened on the right, in the rhythm of the line frequency. The frame scanning current of the blue CRT must therefore have a correction current of the following description superimposed upon it.

(i) While the beam is at the top of the screen the correction current must go from negative to positive as the beam swings from left to right.
(ii) When the beam is half way down the screen the correction current must be zero.
(iii) While the beam is at the bottom of the screen the correction current must go from positive to negative as the beam swings from left to right.

(A "positive" correction current is here taken to mean a deflection towards the top of the projected image.)

The correction current must therefore have the shape drawn in *fig. 8a*, and it might be obtained by multiplying a sawtooth of frame frequency (fig. 8*b*) by a sawtooth of line frequency (fig. 8*c*); how it is actually done will be described later.

At the same time the red CRT must also be supplied with a correction current of the same magnitude but of opposite direction. For the corrections to the vertical scanning for both red and blue tubes, therefore, it will again suffice to generate a single correction current.

It should be observed for the sake of completeness that the addition of correction currents is not the only way of modifying the rasters of the outside tubes; it is possible in theory to give the scanning coils a shape different from the normal. But this involves two complications: first, electron-optical

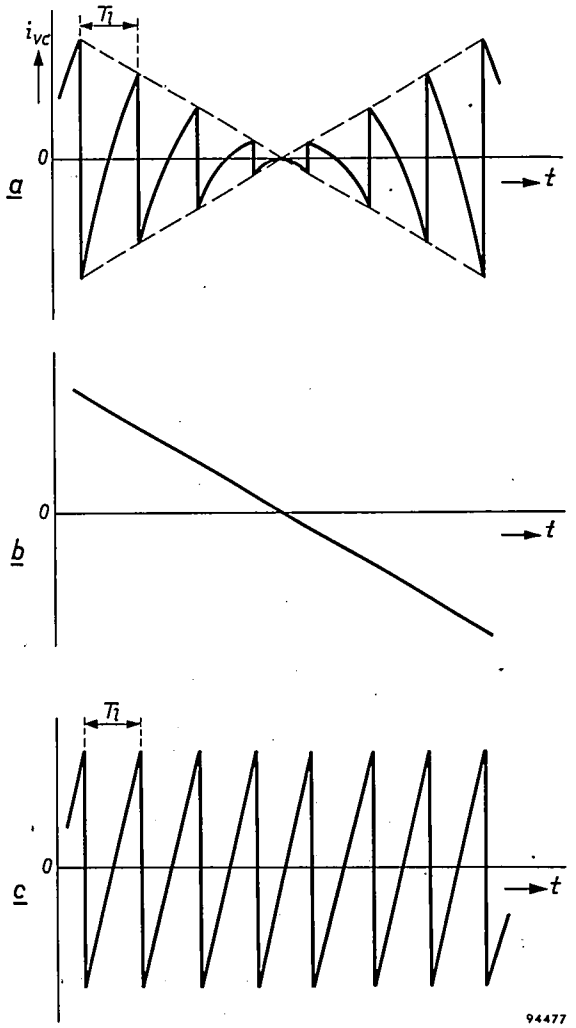


Fig. 8. Trapezium correction. *a*) Waveform of correction current i_{vc} that has to be added to the frame scanning current of the right-hand projector (*B* in fig. 5*a*); the left-hand projector requires correction with a current of the same magnitude but of opposite sense. T_l is the line period. The function drawn in diagram (*a*) is the product of the sawtooth (*b*) of frame frequency, and the sawtooth (*c*) of line frequency.

errors may be introduced and, secondly, the coils in question are difficult to manufacture reproducibly. The method involving special scanning coils is therefore less attractive than that utilizing correction currents.

Optical system of the equipment

The dimensions of the optical systems and projection tubes are determined by what is required of the projected picture as regards size and peak luminance, and also by the luminous efficiency of the phosphors used for the tube screens. Phosphors capable of taking the very high loads that occur and with colour points close to the primary colours laid down in the N.T.S.C. (U.S.A. National Television System Committee) system are listed in *Table I*. Luminous efficiencies in lumens per watt are also indicated.

Table I. Data on the phosphors used.

Primary colour	Composition of phosphor	Coordinates of the colour point		Luminous efficiency lm/W
		x	y	
Red	$(Zn,Be)_2SiO_4$-Mn (with Wratten filter 25)	0.670	0.330	$\eta_r = 2.5$
Green	Zn_2SiO_4-Mn	0.195	0.720	$\eta_g = 21$
Blue	$(Ca,Mg)SiO_3$-Ti	0.160	0.125	$\eta_b = 2.5$

Projection equipment without dichroic mirrors

For equipment not employing dichroic mirrors, the colour points of the phosphors are the same as those of the primary colours forming the projected picture. In order to obtain standard white C (see [1]), fig. 8) from a combination of these primary colours, the luminance values of the primary colours have to be in the following proportions:

$$L_r : L_g : L_b = 29 : 51 : 20.$$

Accordingly, the projection tube screens must dissipate powers in the proportions:

$$P_r : P_g : P_b = \frac{L_r}{\eta_r} : \frac{L_g}{\eta_g} : \frac{L_b}{\eta_b} = 53 : 11 : 36.$$

We see that the red tube screen has to dissipate the highest power. In a normal colour-television picture, white features are responsible for the heaviest loading of all three of the tubes; hence the maximum loading the red tube can be allowed to undergo sets a limit to the maximum brightness that can be obtained in the projected picture. If therefore the three projectors are identical in design, the design of the apparatus as a whole will have to be based on the luminance that the red light contributes to the brightest white parts of the projected picture. If this is desired to have a highlight luminance of L'_{max}, the red contribution will be

$$L_r'_{max} = \frac{L_r}{L_r + L_g + L_b} L'_{max} = 0.29\, L'_{max}. \quad (1)$$

It has been found that the mean luminance of an average television picture amounts to about a third of the peak value. The red tube must therefore supply a total luminous flux of

$$\Phi_r = \frac{\pi b' h'}{\tau_1 \tau_2 a^2 \mu} \frac{L_r'_{max}}{3} \text{ lumen}, \quad (2)$$

where

b' is the width and h' the height (in metres) of the picture on the projection screen;
a the numerical aperture of the projector;
μ the "gain" given by the projection screen (due to directionally selective reflection);
τ_1 the masking coefficient (masking effect of CRT in optical system);
τ_2 the optical transmission coefficient of the system;
$L_r'_{max}$ the maximum luminance in cd/m² of the red image on the projection screen.

The electrical power that the red tube must dissipate in order to supply the above value of luminous flux is

$$P_r = \frac{\Phi_r}{\eta_r}. \quad (3)$$

If p is the maximum power that can be dissipated per unit area of the tube screen, the area bh of the image on the latter must satisfy

$$b\,h = \frac{P_r}{p}. \quad (4)$$

The aspect ratio $b : h$ is laid down at 4 : 3, and this allows us to work out b, h and the diagonal of the picture which should not, of course, exceed the diameter of the fluorescent screen. This gives us the magnification that has to be effected in the optical system of the projector:

$$N = \frac{b'}{b}.$$

Straightforward geometrical-optical reasoning shows that the following relationships must exist between the magnification N, the distances x and x' of object and image from the focus F', the focal length f' and the distance Δ' between the corrector plate and the screen (see *fig. 9*):

$$|N| = \left|\frac{x'}{f'}\right| = \left|\frac{f'}{x}\right| = 1 + \frac{\Delta'}{f'}. \quad (5)$$

For a given magnification, f' would have to be

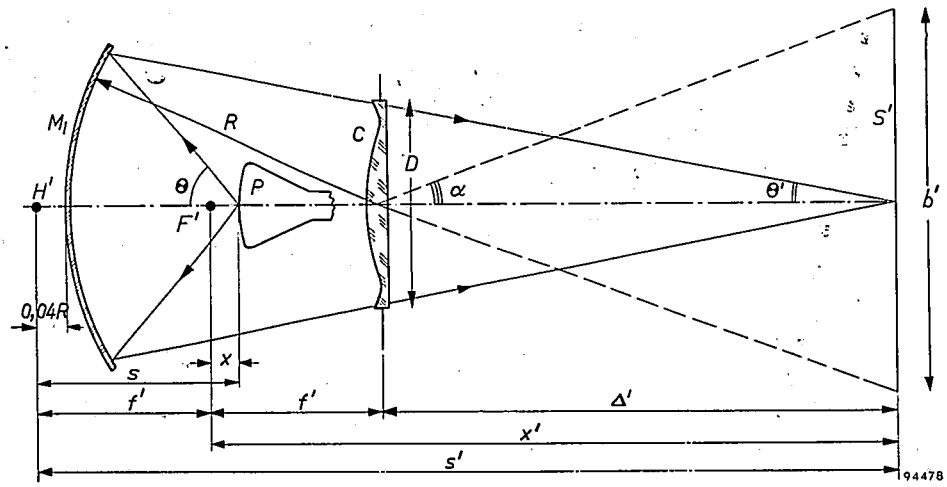


Fig. 9. Schmidt optical system of the "in-line" type. P projection tube. M_1 concave mirror of radius R. C corrector plate of diameter D. S projection screen, the image on which is of breadth b'. F' focus of mirror plus corrector plate. H' pole of mirror plus corrector plate.

increased in direct proportion to Δ', and so also would all the principal dimensions of the optical system. The distance Δ' should therefore be kept as short as possible (this is why the apparatus is set up in front of the viewers). However, there is also a lower limit to Δ', since, as the picture angle $2a$ increases (this being the angle subtended by the width of the projected picture at the centre of the corrector plate — see fig. 9), optical aberrations increase and, apart from that, an increasing fall-off in brightness between the centre and edges of the screen picture becomes evident. There is therefore a certain picture angle $2a_{max}$ at which these effects are still tolerable and which must not be exceeded. Accordingly, the minimum distance between the corrector plate and projection screen is then

$$\Delta'_{min} = \frac{\frac{1}{2}b'}{\tan a_{max}} \quad \ldots\ldots \quad (6)$$

We can now obtain the focal length from (5) and (6):

$$f' = \frac{\Delta'_{min}}{|N|-1} \quad \ldots\ldots \quad (7)$$

The radius of curvature R of the concave mirror is approximately $2f'$.

The diameter D of the corrector plate is worked out as follows. We see from fig. 9 that $D = 2\Delta' \tan \Theta'$. Furthermore, $\sin \Theta' = (\sin \Theta)/|N| = a/|N|$. The angle Θ' is so small that $\tan \Theta' \approx \sin \Theta'$. Hence,

$$D \approx \frac{2a}{|N|}\Delta' \quad \ldots\ldots \quad (8)$$

By way of illustration, a design study will now be outlined for an equipment to project a coloured television picture measuring 2.25 m × 3 m and having a maximum luminance of 20 cd/m². We assume that the picture is projected onto a "beaded" screen having a "gain" of $\mu = 2.8$. A practical value for the numerical aperture is $a = 0.7$. The product $\tau_1\tau_2$ can be assumed to be 0.5 and we can take p to be 4000 W/m². Formula (1) gives the maximum luminance of the projected red image:

$$L_r'_{max} = 0.29 \times 20 = 5.8 \text{ cd/m}^2.$$

The following values are obtained successively from (2), (3) and (4):

$$\Phi_r = 59 \text{ lm},$$

$$P_r = \frac{59}{2.5} = 24 \text{ W},$$

$$b\,h = \frac{24}{4000} = 6 \times 10^{-3} \text{ m}^2.$$

From these we work out the dimensions of the image on the CRT screen:

$$b = 9 \text{ cm}; \; h = 6.7 \text{ cm}; \text{ diagonal} = 11.2 \text{ cm}.$$

A projection tube with dimensions of the MW 13-38 (which has a screen of diameter 13 cm) can therefore be used.

The magnification N is equal to $b'/b = 300/9 \approx 33$. In a Schmidt optical system the maximum picture angle $2a_{max}$ is round about 30°, and on this basis we can determine the minimum distance between corrector plate and projected picture from (6):

$$\Delta'_{min} = \frac{\frac{1}{2} \times 3}{\tan 15°} = 5.60 \text{ m}.$$

Formula (7) gives the focal length:

$$f' = \frac{560}{33-1} = 17.5 \text{ cm}.$$

The concave mirror must therefore have a radius of curvature of $R \approx 2f' = 35$ cm. The periphery of the mirror is usually given a diameter approximately equal to R; with a smaller mirror, differences of illumination over the projection screen area would be too apparent.

Finally, the diameter of the corrector plate is obtained from (8):

$$D \approx \frac{2\times 0.7}{33} \times 560 \approx 24 \text{ cm}.$$

Projection equipment embodying dichroic mirrors

The foregoing applies to an equipment that does not embody dichroic mirrors. The design procedure for a projector that does employ such mirrors is similar, except that the design must be based on a different set of relationships between L_r, L_g and L_b, the reason being that the selective transmission and reflection of the dichroic mirrors causes the colour points of the primary colours to shift. In addition, transmission losses in the mirrors must be taken into account in the calculation of the electrical powers that must be dissipated by the projection tubes. As already stated, the "folded" optical system is preferred where dichroic mirrors are to be used; but such a system has a lower transmission coefficient. If the projectors are to have the same dimensions as in the foregoing, the overall result of all this is that highlight luminance of the projected picture will be reduced by roughly 50%. Projectors designed on the above lines and used in combination with dichroic mirrors will therefore be able to produce a maximum luminance of about 10 cd/m² on the projection screen.

Table II gives the dimensions of the four equipments built in this laboratory. Column 1 relates to a projector corresponding roughly to the example worked out above.

Mechanical details of the optical systems

The projection tube, complete with focussing and deflection coils, must be mounted in its optical system in such a way as to enable *optical focussing* to be carried out by altering the position of the tube with respect to the concave mirror. For this purpose it must be possible to move the tube forward and backward along its own axis and also to turn it about two axes perpendicular to each other and to the axis of the tube, running through the centre of the tube screen. *Fig. 10* shows how this is done in the projector depicted in fig. 4. The projection tube P, together with coils D and F, is suspended in a yoke B consisting of a flanged plate holding the tube assembly. Two threaded holes in the plate carry threaded rods I, whose ends rest in recesses U and V, being retained by tension springs between the plate and the chassis. The third point of support is provided by an angle bracket fixed to the plate, resting on a bolt (W) threaded in a chassis member. The points U and V lie on the horizontal line passing through the centre of the tube screen. If the adjusting screws marked I are turned to the same extent in the same direction, the tube moves forward or back; if they are turned to the same extent in opposite directions, it turns about the vertical axis through the centre of its screen. Adjusting the knob II on W causes the tube to turn in a vertical plane, about the axis UV. These operations, each of which is substantially independent of the others, allow the whole area of the projected image to be brought into focus.

Table II. Data on the Schmidt optical systems used in the four colour-television projectors constructed.

	1	2	3	4
Dichroic mirrors	no	yes	yes	yes
Type of system	"in-line"	"folded"	"folded"	"folded"
Picture on projection screen:				
Height h' **)	2.25	2.25	0.35	0.22
Breadth b' **)	3.00	3.00	0.46	0.29
Max. luminance L'_{max} (cd/m²)	20	10	200	200
Contrast ratio	35	25	25	25
Definition *)	600	425	450	500
Screen "gain" μ	2.8	2.8	7	7
Diam. of tube screen **)	0.13	0.13	0.06	0.06
Radius of curvature R of concave mirror **)	0.40	0.40	0.20	0.20
Focal length f' **)	0.208	0.208	0.104	0.104
Diameter D of corrector plate **)	0.28	0.25	0.125	0.125
Projection distance Δ' **)	6	6	0.825	0.645
Magnification $\|N\|$	31	31	9.2	7.4

*) Number of lines; determined with an RMA test plate.
**) In metres.

Vignetting is an important point in the design of optical systems. The endeavour must be to keep the amount of vignetting the same to the left and right, so that both halves of the projected image are equally bright. In an apparatus embodying dichroic mirrors, the images on the three tube screens are in different orientations with respect to their optical systems, owing to the various mirror inversions; if therefore local colour variations are to be avoided, it is particularly important that the vignetting of the optical systems should be the same at the top and bottom of the picture.

We shall deal later with the arrangements for altering the position of the coils. It may be merely

observed at this point that interaction between the focussing and scanning coils may lead to an error in superposition that cannot be corrected; this error takes the form of differences in curvature between horizontal and vertical lines in the three primary-colour images. Such interaction must therefore be obviated, one way of doing so being to fit a copper ring and a ring of material that is a good magnetic conductor between the focussing and the deflection coils.

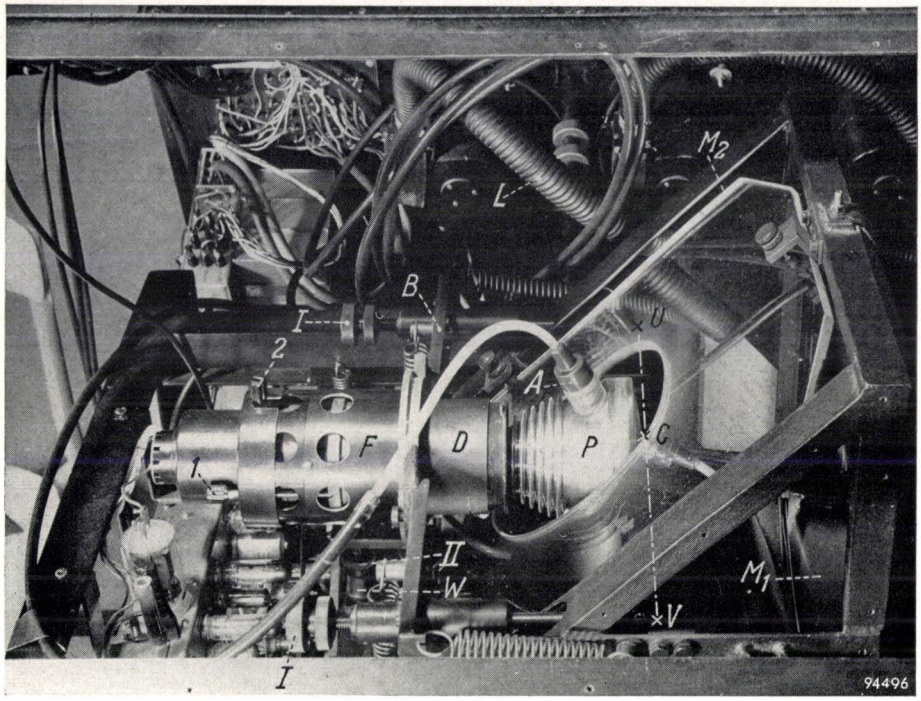


Fig. 10. One of the primary-colour projectors of the large-screen apparatus shown in fig. 4. P is the projection tube, A its anode connection, D the scanning coils and F the focussing coil. M_1 is the concave mirror, M_2 a plane mirror inclined at 45°. L is a flexible pipe through which cooling air is blown against the window of the tube. Part of the video amplifier is visible below, on the left. Tube and coils are suspended in a yoke B supported at points U, V and W. By turning the screws I-I both in the same direction the tube is moved backward or forward along its own axis; turning them in opposite directions causes the tube to swing laterally about the vertical line through the centre C of its fluorescent screen. Turning screw II causes the tube to turn about the horizontal line UV passing through C. Controls *1* and *2* are for moving the focussing coil (see below).

Circuits and adjustments for obtaining superposition of the primary-colour images

If the three primary-colour images are to be brought into correct superposition on the projection screen, the apparatus must allow of a number of adjustments and corrections, some of which have already been mentioned.

1) In an apparatus with converging projectors (fig. 5) *trapezium correction* (figs. 6 and 8) will be necessary to turn the two images with that shape into rectangular ones.

2) The three reactangular images must be of the same size; if must therefore be possible to control the *amplitude* of the scanning currents.

3) It may be necessary to correct the *linearity* of the scanning, in order to ensure movement of the three spots across the projection screen at the same speed.

4) *Displacements* of the three images will be necessary, in order to make their centres coincide.

5) *Rotational alignment* of the images will be necessary, in order to bring their longer edges into parallelism.

In addition to all this it will be necessary to take into account the *spread* in the electrical and mechanical properties of certain components: there will be differences in the deflection sensitivity of the projection tubes, in the position of the electron gun in relation to the tube axis, in the accuracy with which the angle between the two scanning directions approximates to 90° for the different sets of coils, and so on.

Most of the corrections can be incorporated in the scanning arrangements, and we shall therefore start by discussing the appropriate circuits.

The scanning circuits.

Apart from the necessary corrections the deflection of the electron beams of the three projection tubes must be identical. Hence, if the three scanning units were exactly similar and if the three tubes had exactly the same deflection sensitivity, the scanning currents would likewise have to be exactly equal for each of the two scanning directions. In order to ensure from the outset that the said currents will be as nearly equal as possible, it is advisable to take them from a common generator for each scanning direction, rather than separate ones for each CRT. With separate sawtooth generators, the free oscillations that might arise in the deflection coils and transformer windings would be independent of one another, and this would render the exact superposition of the three primary-colour images impossible. Moreover, changes in the properties of the scanning generator tubes would, in the long run, give rise to discrepancies in the scanning current amplitudes, and hence to dimensional differences in the primary-colour images. Furthermore, such a method would be expensive.

The common scanning generator can be made to supply the three deflection coils connected either in series or in parallel. It might be thought that, with the series arrangement, exactly the same current would flow through the three coils. However, each coil has a stray capacitance in parallel with it, so that the arrangement in fact constitutes three oscillatory circuits in series, each able to oscillate at its natural frequency. In general, the three frequencies will not be the same and discrepancies will again arise in the scanning currents, resulting in superposition errors. Another objection is that two of the three coils carry high peak voltages with respect to earth, which might give rise to difficulties.

If the coils are in parallel, the same voltage is across all three and hence it is impossible for them to oscillate independently of one another. By inserting a variable impedance in series with each coil, the amplitude and linearity of the current through it can be adjusted to some extent. It is very easy to introduce other corrections that may be necessary, as will presently be evident. For these reasons it is only the parallel arrangement that is used.

Skew-correction

A scanning unit should deflect the beam in two directions with an angle of exactly 90° between them. In practice there is a certain amount of spread in this angle as exhibited by different deflection yokes. Any deviation from 90° manifests itself on the receiver screen as a departure from exact perpendicularity between the vertical and horizontal lines of the picture. The deviation is so slight in normal units that its effect is unnoticeable in monochrome television. In colour television, on the other hand, the deviations may be troublesome because they may differ in magnitude and direction as between the three primary-coulour images. To obviate them, either scanning units must be used that have been built with extreme care, or the fault must be corrected by electrical methods.

In order to correct skewness in the projected image electrically, the line scanning coils are injected with a correction current that is proportional to i_V, the frame scanning current. Since the frequency of i_V is so low that the scanning coils respond to i_V practically as resistances, it will suffice to introduce variable and reversible voltages, proportional to i_V, in series with the coils. These voltages should preferably be obtained from separate sources, in order to prevent the adjustments affecting one another. This skew correction only needs to be provided for in two of the three primary-colour projectors.

Adjustment of amplitude

The output circuit of the common sawtooth generator for the frame scanning is as shown in *fig. 11*. As already stated the frame scanning coils may be regarded as behaving as resistances with regard to i_V. They also possess some inductance, however, and the linearity of the sawtooth current would not be preserved if its amplitude were adjusted with the aid of a variable resistor alone, placed in series with the coil. The ratio of the resistance of the combination to its inductance would vary, and so would the shape of the current, causing errors of superposition. If therefore the shape of the sawtooth supplied to the three CRTs is to be kept uniform, a small variable inductance must be placed in series with each coil (fig. 11).

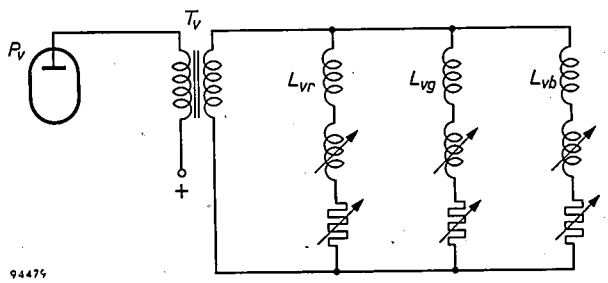


Fig. 11. Adjustment of the amplitude of the frame scanning currents. L_{Vr}, L_{Vg} and L_{Vb} are the frame scanning coils of the red, green and blue tubes respectively. P_V is the output tube of the common sawtooth generator. In series with each of the coils are a variable resistor and a variable inductance, the purpose of the latter being to keep the ratio between total resistance and inductance constant.

With the line scanning coils it is otherwise; these behave more or less as inductances with regard to i_h, the current flowing through them. Consequently, the adjustment of amplitude must be effected by means of a variable inductance in series with the coil, a small variable resistance also being present, so that the ratio of resistance to total inductance can be kept constant. It is desirable that the loading of the horizontal time-base should be kept as nearly constant as possible; accordingly, a potentiometer frequency is fed in via pentode P_{lin} and transformer T_{lin}; it is controlled by means of the cathode resistor of P_{lin}. This allows for the compensation of any lack of linearity in the current from the common sawtooth generator, and is effective for the line scanning of all three projection tubes.

The coarse adjustment of scanning amplitude is carried out elsewhere in the line scanning generator, one control serving for all three scanning coils. The same applies to the frame scanning generator.

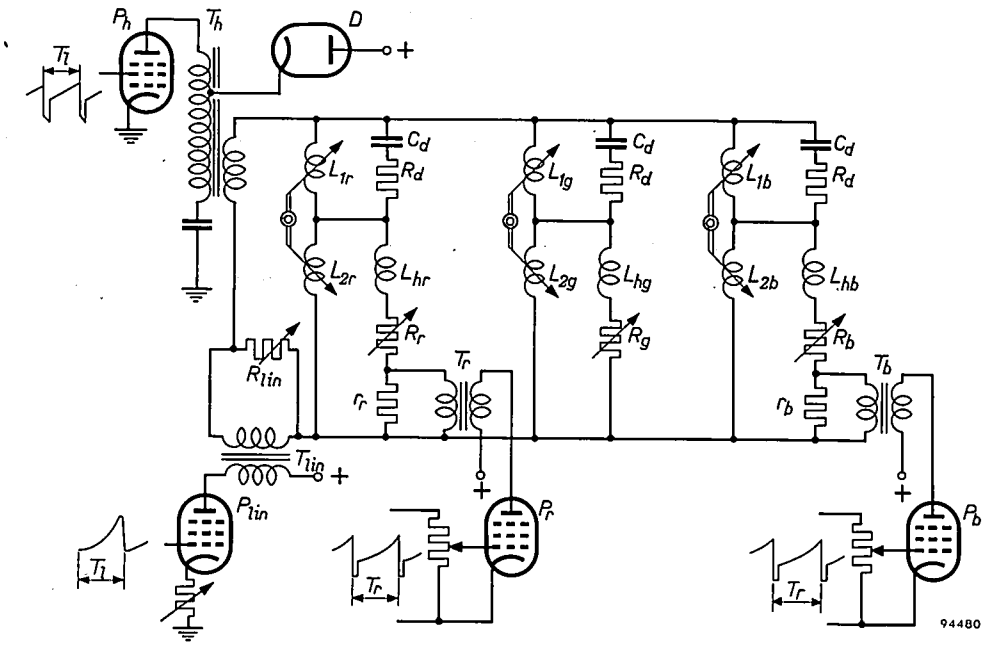


Fig. 12. Adjustment of the amplitude of the line scanning currents. L_{hr}, L_{hg} and L_{hb} are the line scanning coils of the red, green and blue tubes respectively. P_h is the output tube of the common sawtooth generator, T_h transformer. D series efficiency diode (usual in line time-bases). L_{1r}-L_{2r}, L_{1g}-L_{2g} and L_{1b}-L_{2b} are potentiometrically connected variable inductances. R_r, R_g and R_b are variable resistances allowing the ratio between total inductance and resistance to be kept constant. The R_d-C_d networks are for damping free oscillations. The sawtooth current delivered by T_h is rendered linear via pentode P_{lin}, transformer T_{lin} and resistor R_{lin}, the linearity control being the cathode resistor of P_{lin}. A correction current proportional to i_v is fed to the red and blue CRTs via P_r-T_r-r_r and P_b-T_b-r_b respectively (for skew-correction). T_l = line period. T_r = frame period.

arrangement of inductances, as shown in *fig. 12* is to be preferred.

It may also be seen from fig. 12 how the skew correction is effected in a large-screen projector. Transformers T_r and T_b are fed, independently of each other, from pentodes P_r and P_b respectively, these tubes being controlled by an adjustable fraction of the sawtooth voltage on the grid of the frame-scan output tube. Resistors r_r and r_b thus have voltages across them which have the same shape as i_v, and which, by correcting the current i_h flowing through line-scanning coils L_{hr} and L_{hb}, serve to correct skewness in the projected images.

A further voltage of sawtooth shape and line

Trapezium correction

In apparatus with the arrangement of projectors shown in fig. 5, the frame and line scanning of the two outer CRTs must be made subject to correction, as already pointed out (figs. 6, 7 and 8).

A circuit for superimposing a correction current on the *frame* scanning current appears in *fig. 13*. The uncorrected sawtooth current arrives via transformer T_v, the correction current via transformer T_{vc}. The circuit is designed in such a way as to maintain the correct relative direction of the two currents and to prevent mutual interference between the generators producing them.

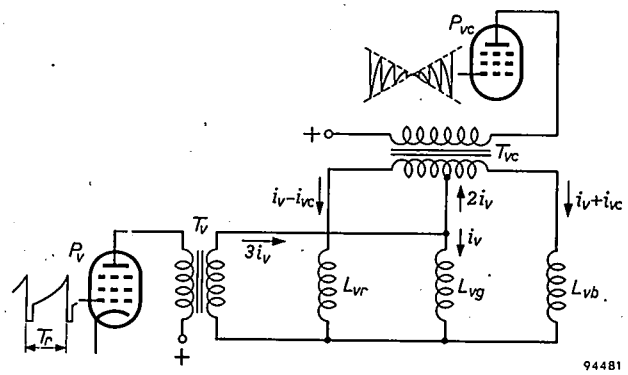


Fig. 13. Circuit for adding i_{vc}, the current for correcting trapezium distortion, to the frame scanning current i_v in the inductances L_{vr} and L_{vb}. The correction current is delivered by tube P_{vc}. T_{vc} is a centre-tapped transformer. The remaining letters have the same meaning as in fig. 11.

Fig. 14 shows the essentials of a circuit from which the correction current can be obtained. Voltages e_1 and e_2 are sawtooth waveforms of opposite sense and have the frame frequency, $1/T_r$. Switches S_1 and S_2 (in reality they are valves) are closed only during the flyback of the line sawtooth. When they are open, good approximations for the voltages u_1 and u_2 across capacitors C_1 and C_2, are:

$$u_1 \approx (E + e_1)\,\frac{t - nT_l}{RC},$$

$$u_2 \approx (E + e_2)\,\frac{t - nT_l}{RC},$$

where E is a direct voltage, R is the value of the resistors in series with each capacitor (both capacitors have the value C), T_l is the line period, n is an integer indicating the order of the line in the raster ($1 \leqslant n \leqslant 625$ for the 625-line system) and t is an instant such that

$$nT_l < t < (n+1)T_l.$$

If $\hat{e}$ is the peak value of e_1 and e_2, and $\beta = \hat{e}/E$, then

$$E + e_1 = E\left(1 - \beta + \beta\,\frac{t}{T_r}\right)$$

and

$$E + e_2 = E\left(1 - \beta\,\frac{t}{T_r}\right).$$

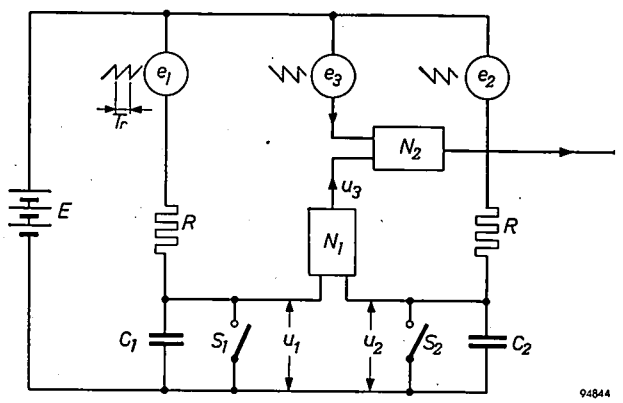


Fig. 14. Schematic diagram of the circuit for generating i_{vc}, the current for correcting trapezium distortion. Switches S_1 and S_2 (in reality they are valves) are closed only during the flyback of the line sawtooth. e_1, e_2 and e_3 are sawtooth voltages having the frame frequency. The subtraction $u_1 - u_2 = u_3$ takes place in network N_1, the addition $u_3 + e_3$ in network N_2.

This gives us the following for the voltages across the capacitors:

$$u_1 \approx E\left(1 - \beta + \beta\,\frac{t}{T_r}\right)\frac{t - nT_l}{RC},$$

$$u_2 \approx E\left(1 - \beta\,\frac{t}{T_r}\right)\frac{t - nT_l}{RC}.$$

These voltages have the waveforms shown in *figs. 15a* and 15*b*, respectively. The difference between the two is a voltage u_3 whose waveform is as shown in fig. 15*c*:

$$u_3 = u_1 - u_2 \approx \beta E\left(-1 + \frac{2t}{T_r}\right)\frac{t - nT_l}{RC}.$$

This needs only the addition of a sawtooth voltage

$$e_3 = \tfrac{1}{2}\beta E\left(1 - \frac{2t}{T_r}\right)\frac{T_l}{RC}$$

at the frame frequency (fig. 15*d*) to give a voltage of the desired shape (fig. 15*e*, cf. fig. 8*a*). The subtraction of u_2 from u_1 takes place in network N_1, the addition of e_3 to u_3 in network N_2.

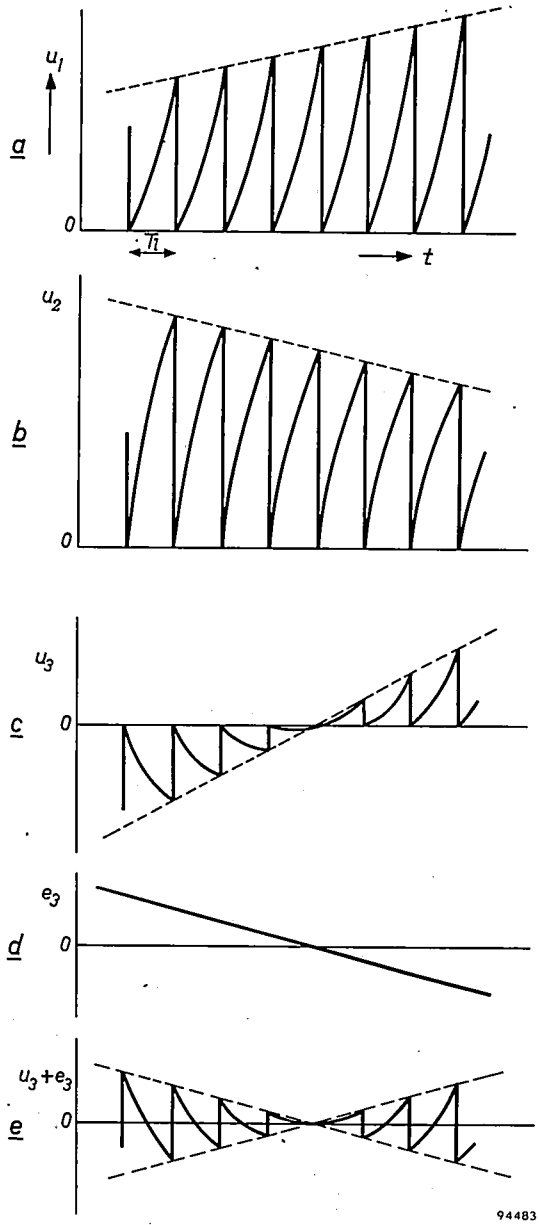


Fig. 15. Voltage waveforms u_1 and u_2, $u_3 (= u_1 - u_2)$, e_3 and $u_3 + e_3$ occurring in the circuit of fig. 14.

The *line* scanning currents of the outer CRTs must be corrected in opposite senses by currents such as were indicated in fig. 6*b*. The shape is approximately parabolic. A correction current of this kind can be obtained by introducing a sawtooth voltage e_{hc} of line frequency in series with the scanning coils. Over one line period e_{hc} can be written as

$$e_h = A\,t \quad (\text{for } -\tfrac{1}{2}T_l \leqq t \leqq +\tfrac{1}{2}T_l),$$

A being a factor of proportionality. In the parallel arrangement of coils of inductance L, this voltage produces a current i_{hc} which, if losses can be neglected, may be written as:

$$i_{hc} = \frac{1}{L}\int e_{hc}\,dt.$$

Apart from a constant of integration, this is equal to $(A/2L)t^2$ and thus corresponds to a curve of parabolic shape. The circuit is given in *fig. 16*. Like that of fig. 13, it is balanced in such a way as to prevent interaction between the two generators.

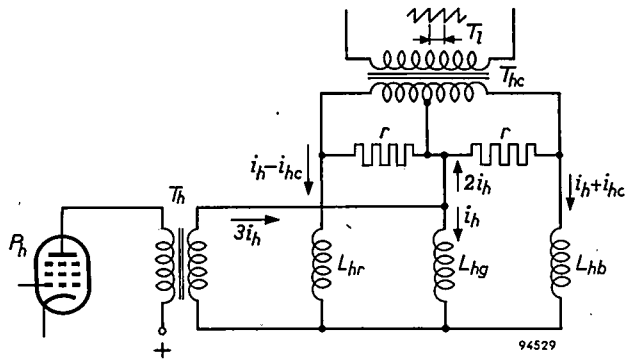


Fig. 16. Circuit for adding correction current i_{hc} to the line scanning current i_h in the inductances L_{hr} and L_{hb}. Transformer T_h, supplied from the output tube P_h, delivers an uncorrected sawtooth current $3i_h$. The centre-tapped transformer T_{hc} supplies resistors r with a sawtooth voltage, which causes i_{hc}, a correction current of parabolic shape, to flow through inductances L_{hr} and L_{hb}.

Displacement and rotational alignment of the images on the projection screen

If the three primary-colour images are to be in exact registration, it must be possible to shift them across the projection screen until their centres coincide, and then to turn them about their common centre until the lines of each raster are exactly horizontal. The latter operation is carried out by turning the deflection coils about the tube axes by means of trimming knobs.

Displacement of the images can be effected in various ways; it can be done with the aid of adjustable direct currents, by turning the focussing coils about axes perpendicular to the axis of the tube, or by altering the alignment of the primary-colour projectors.

1) If an adjustable direct current be passed through each of the six scanning coils, a steady magnetic field of adjustable strength and direction will be set up across each of the CRTs; by this means each electron beam can be biased in direction to the desired degree. The advantage of this method is that the operation can be performed by remote control. But it has a number of disadvantages; one is that the currents have to be stabilized, owing to the temperature dependence of the resistance of the scanning coils.

If the scanning currents are already being subjected to various kinds of correction, there are objections to passing an adjustable direct current through the coils in addition to the others. Separate windings can be incorporated in the coils for carrying these direct currents, but then a high impedance has to be placed in series with the extra windings, since they would otherwise act as short-circuited loops coupled to the scanning windings.

Fitting the direct-current coils in a position before the focussing coils (i.e. nearer the electron gun) is another possibility. Care has to be taken that they do not interfere with the focussing of the beam, and cause part of it to strike the tube diaphragm.

2) It is also possible to give a directional bias of adjustable strength and direction to the electron beam by altering the position of the focussing coil with respect to the axes of the tube. This calls for a mechanical contrivance enabling each of the focussing coils to be turned about two axes at right angles to each other and to the axis of the tube. The method has proved to be very stable and quite satisfactory in practice, in spite of one objection. This arises from slight differences in the position of the electron gun in different projection tubes. For correct superposition of the projected images, it is necessary that the three focussing coils should occupy different positions with respect to the nominal axis of the system. If the "general focussing control" is turned once the projected images have been made to coincide, the three focussing currents will increase or decrease to the same extent, but the images will shift arbitrarily (in directions depending on the position that the gun happens to have in each tube) and superposition will be lost. It is therefore necessary to start by focussing the three images, each of the three focussing currents being separately adjusted, and to adjust for superposition only after this has been done.

The difficulty might be avoided by first setting the focussing coils into a position such that the images would not shift when the value of the focus-

sing current was changed. However, owing to the position of the electron gun, this would generally result in part of the CRT image falling outside the tube screen, which would obviously be quite impermissible. In any case it is desirable that the CRT images should as far as possible occupy the same position on the tube screens, to obviate differences in distortion (due to both optical and electron-optical causes).

3) Altering the alignment of the primary-colour projectors is a purely mechanical operation which can easily be made very stable. For this purpose it must be possible to turn each projector about two axes perpendicular to each other and to the axis of the projector itself. Preferably the two axes should pass through the centre of the corrector plate, but this need not necessarily be the case, if using other axes simplifies the mechanical design. Movement of the projectors about the two axes will then involve a change in their distance from the projection screen, but so small a change that it will not cause any noticeable error.

In the projectors constructed in this laboratory use has been made of a combination of methods (2) and (3); the CRT images are moved into as nearly central a position as possible on the tube screens by turning the focussing coils (i.e. by turning screws *1* and *2*, see fig. 10). The three projected images are then made to coincide on the viewing screen by altering the alignment of the projectors.

Extra-high-tension supply

Our small-screen projection sets are supplied with EHT at 25 kV, the large-screen sets with EHT at 50 kV. It is very important that the EHT supply should be kept strictly constant in spite of fluctuations in the load and in mains voltage. Changes in EHT voltages are accompanied by the following phenomena:

a) A change in the size of the images, due to the fact that the deflection sensitivity depends on the EHT value. Small changes of dimension are permissible providing they affect the three primary-colour images simultaneously and to the same extent. In view of this, the EHT supply for the three projection tubes is provided from a common source.

b) Mutual displacement of the three primary-colour images. The directional bias imposed on the electron beam referred to in the previous section generally differs in amount and direction in the three projection tubes, being dependent on the position that the electron gun happens to have; the consequence is that the three primary-colour images move in different directions when the EHT value changes, superposition thus being lost.

c) Loss of electron-optical focus, resulting in loss of definition.

The EHT value can be rendered more or less independent of the load current by applying parallel stabilization and/or negative feedback.

The term "parallel stabilization" refers to a circuit in which a triode in parallel with the load is controlled in such a way that the total load, consisting of the triode and the three projection tubes, is kept constant. The method is suitable for all systems of generating EHT. Triodes have been developed for this purpose that are capable of standing voltages as high as 80 kV, allowing parallel stabilization to be applied in all cases arising in practice.

Another method makes use of a combination of current and voltage feedback into the EHT generator, which should preferably consist of an oscillator generating a voltage with a fairly high frequency plus a cascade arrangement of diodes and capacitors (voltage multiplier) [6]). A control voltage whose value is dependent both on the load current and on the value of the EHT acts on the oscillator — being applied to its control grid, for example — and on the control curcuit of the regulated HT supply of the oscillator. In this manner it is possible to keep the EHT value constant to within 0.5% over the whole range of load currents.

Neither stabilizing arrangement reacts quickly enough to sudden changes in the load. The circuit should include a buffer capacitor of sufficiently high value to absorb any rapid changes.

The effect of fluctuations in mains voltage can largely be neutralized by using a mains voltage stabilizer or by electronic stabilization of the oscillator HT supply.

Proportional brightness control

It is desirable that a colour-television projector should have facilities for adjusting background brightness by means of a single control. This control, called a proportional brightness control, allows the bias voltages on the control grids of the three CRTs to be altered simultaneously, and in such a way that the colour of the projection screen is always a neutral grey when no video signal is coming in, whatever the position of the control knob. The condition for this may be derived as follows.

---

[6]) J. J. P. Valeton, Philips tech. Rev. 14, 21-32, 1952/53.

The luminous flux Φ emitted by a cathode ray tube is related to the grid voltage (measured from the cut-off value) by the following approximate expression:

$$\Phi = c\, V^{\gamma}, \quad \ldots\ldots \quad (9)$$

where c is a constant and the exponent γ has a value between 2 and 2.5. The condition for obtaining standard white C on the projection screen, from a mixture of the primary colours employed, is that

$$\Phi_r : \Phi_g : \Phi_b = 29 : 51 : 20.$$

The voltages on the three control grids must therefore be in the following proportions:

$$V_r : V_g : V_b = \left(\frac{29}{c_r}\right)^{1/\gamma} : \left(\frac{51}{c_g}\right)^{1/\gamma} : \left(\frac{20}{c_b}\right)^{1/\gamma}, \quad (10)$$

c_r, c_g and c_b being values of the constant c in (9), appropriate to the red, green and blue tubes respectively. Since the proportions on the right-hand side of (10) are independent of brightness, it must be possible to achieve the aim in view with linear variable resistors.

The circuit employed is shown in *fig. 17*. The brightness of the red, green and blue tubes is adjusted separately by means of control knobs 2_r, 2_g and 2_b, Each knob actuates two variable resistors (R_r' and R_r'', R_g' and R_g'', R_b' and R_b'') ganged in such a way that their total resistance always remains constant; hence the voltage across potentiometers P_r, P_g and P_b is independent of the position of controls 2_r, 2_g and 2_b. Potentiometers p_r, p_g and p_b are the proportionality controls; they select the fraction of the voltage at the sliders of P_r, P_g and P_b (ganged, and adjusted by knob *1*) that is applied to the grids of the projection tubes. Once 2_r, 2_g and 2_b, and p_r, p_g and p_b, have been correctly set, the background brightness can be adjusted by turning knob *1*, without any change in the neutral tint on the projection screen. The circuit has the attractive characteristic that, when knob *1* is in the zero position, the right settings of 2_r, 2_g and 2_b can be found irrespective of the settings of p_r, p_g and p_b. If this were not so, finding the right settings of knobs *2* and knobs p would involve repeated adjustment of first the one set and then the other.

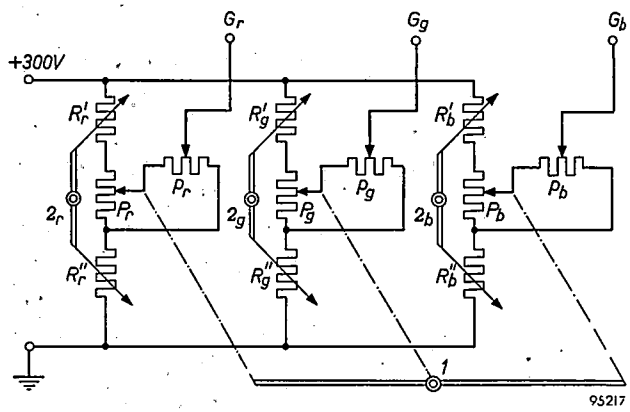


Fig. 17. Schematic diagram of the proportional brightness control. Bias voltages for the red, green and blue CRTs are derived from points G_r, G_g and G_b. *1* is the knob for adjusting general brightness level via the ganged potentiometers P_r, P_g and P_b. 2_r, 2_g and 2_b are knobs for adjusting the brightness of the red, green and blue images separately. The potentiometers p_r, p_g and p_b are for obtaining the right proportions between the brightness values of the three images.

It may be said that the separate brightness controls *2* serve to compensate differences in the cut-off voltages of the projection tubes and that the proportionality controls p serve to compensate differences in slope. In reality the projection tubes do not obey law (9) exactly; however, since it is generally only necessary to adjust background brightness over a small range, departures from the correct colour balance are only very slight.

The video amplifiers

Each of the three projection tubes has its own video amplifier. Besides performing all that is normally required of a video amplifier, the three amplifiers must satisfy special requirements arising from the fact that they are together responsible for the colour rendering of the picture on the projection screen.

In colour television systems, as in black-and-white systems, measures are taken at the transmitter to compensate non-linearity in the receiver picture tube (as exemplified by eq. (9) [7]). In theory, therefore, the receiver video amplifiers should be completely linear, so that the overall channel (camera to projection screen) will be linear, as it must be if colours are to be reproduced correctly. If the three amplifiers are fed with the same video signal, the amplitudes at the three inputs being in certain proportions, a monochrome image must necessarily appear on the screen. However, if the characteristics of the projection tubes differ from the nominal characteristic on which the compensation at the transmitter is based, the range of greys between black and white will no longer be faithfully reproduced. It will in fact contain regions that, instead of being neutral in tone, will exhibit a tint that depends on the differences between the actual and the nominal characteristics of the CRTs. What is more, few of the other colours will be faithfully reproduced.

[7]) See the concluding observations in article I (footnote [1])).

As we have seen earlier, to get a given grey shade on the viewing screen, it is necessary that the quantities of red, green and blue light should be present in a certain specific proportion. Because the characteristics of the CRTs conform to the law $\Phi = cV^{\gamma}$, the control voltages will again be in specific — although different — proportions. But in practice the light flux from a CRT screen does not obey the law exactly, one of the main causes being the saturation of the phosphors. There are differences in the properties of the three phosphors; moreover, the anode currents of the CRTs are not equal when adjusted to produce a grey shade on the projection screen. In consequence, deviations from the law $\Phi = cV^{\gamma}$ are not the same for the three primary colours, being greatest for the red phosphor which exhibits the greatest degree of saturation. The saturation effect is particularly strong in apparatus for large-screen projection, which involves a high specific loading of the phosphors. In practice the control voltages on the three CRTs are adjusted to values such that grey shades are properly reproduced in both the lighter and darker parts of the picture (*fig. 18*); the values chosen for the bias voltages are such that the tubes are just cut off when no signal is coming in. It will be seen from fig. 18

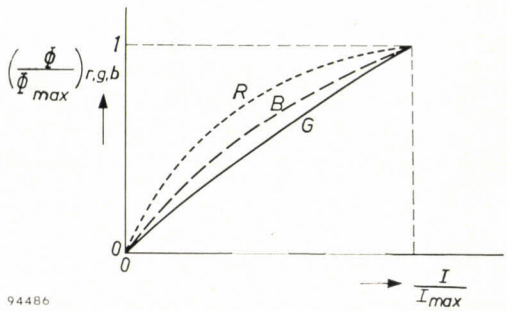
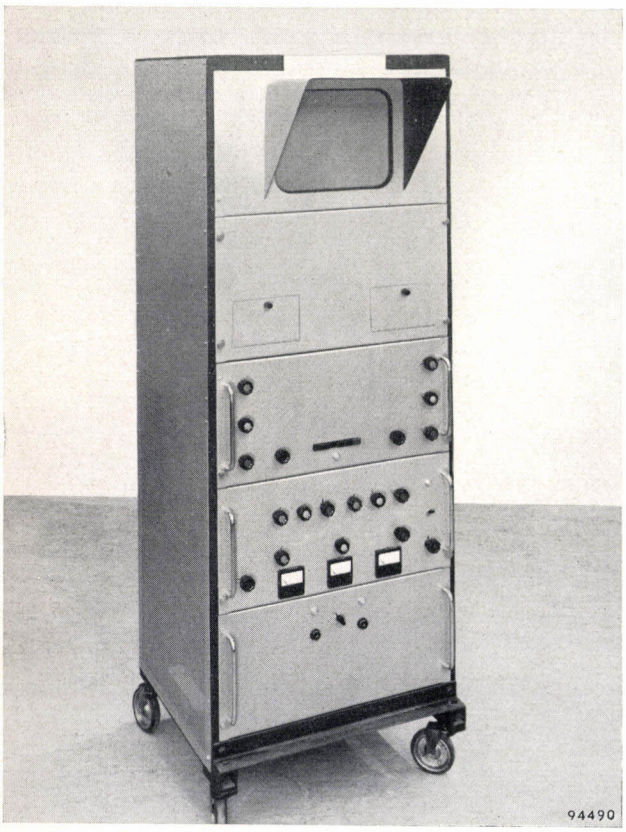
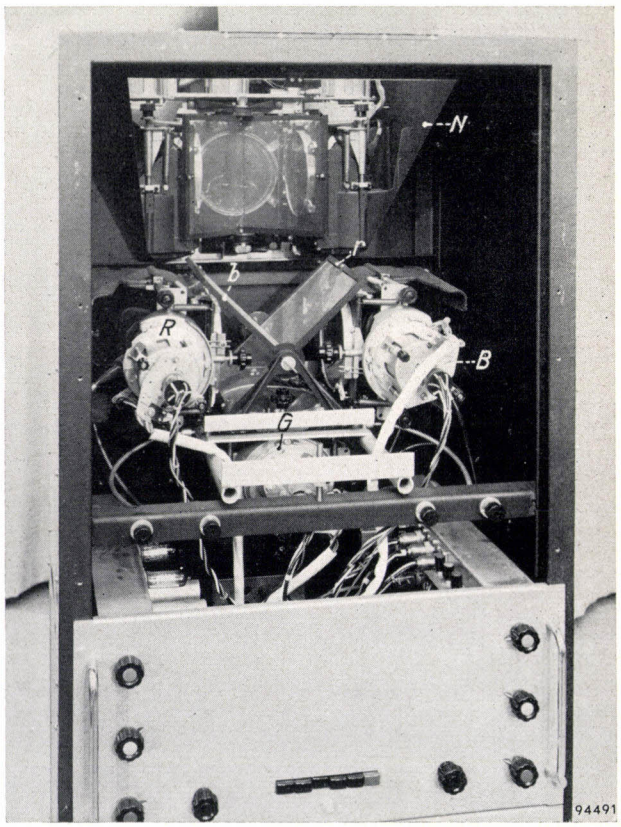


Fig. 18. Ratio of the luminous flux Φ to Φ_{max} as a function of the ratio of anode current I to I_{max}, for the red, green and blue projection tubes (R, G and B respectively). Saturation is heaviest in the red CRT and lightest in the green one.

that the proportion of red light on the projection screen is too great throughout the range of greys between black and brightest white. This effect can be compensated by introducing non-linearities into the video amplifiers such that amplification is higher in the darker parts of the picture than it is in the



a



b

Fig. 19. Colour television projection set for studio monitoring (picture size 22 cm × 29 cm, maximum luminance 200 cd/m²). *a*) The complete set, with transmission projection screen of polyvinyl chloride. *b*) View of the set with the two upper panels removed; R, G and B are the primary-colour projectors, which are on the folded-beam principle. r and b are the crossed dichroic mirrors. N is a plane mirror inclined at 45°.

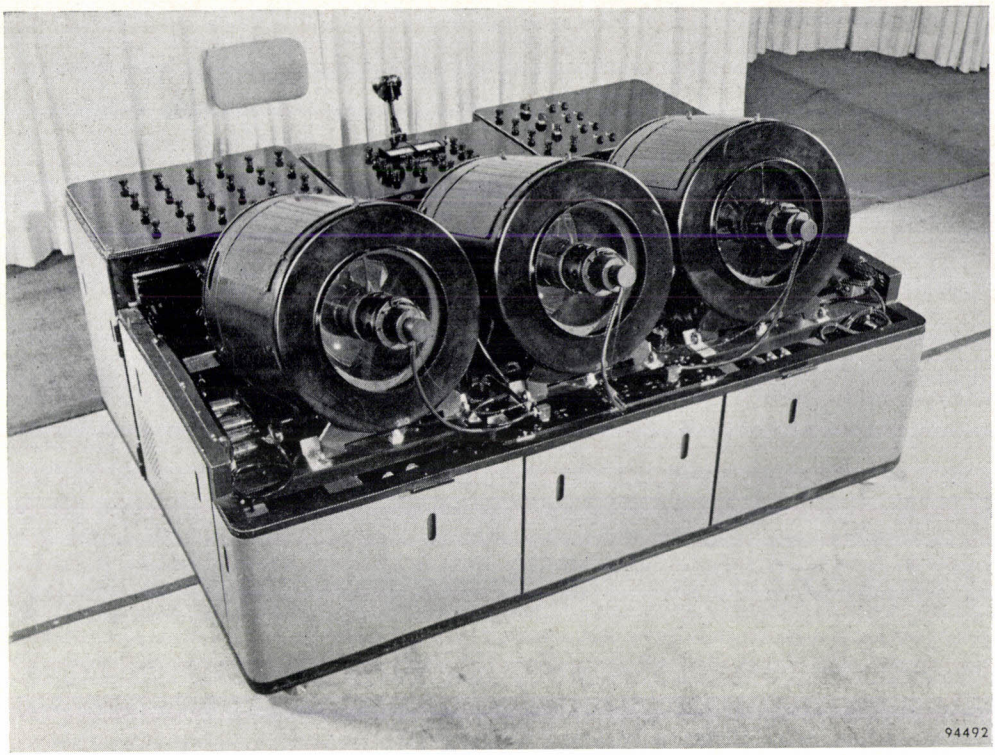


Fig. 20. Large-picture projection set with cover removed. The Schmidt optical systems, which are of the "in-line" type, are mounted side by side. See also figs. 21 and 22.

lighter. This implies a voltage characteristic that can often be approximated by two straight lines of different slope [8]).

In addition, the amount of amplification given by the three amplifiers must remain as nearly constant as possible and be independent of the properties of the electronic tubes with which they are equipped. For this to be so, it is often necessary to apply negative feedback and to stabilize the heater current.

[8]) F. H. J. van der Poel and J. J. P. Valeton, Philips tech. Rev. **15**, 221-232, 1953/54.

Of course, each amplifier must be provided with a contrast control with which the amplification can be adjusted in order to achieve correct colour reproduction.

The projectors constructed

Many of the principles dealt with above have been applied in a number of colour television projectors built in the Philips Eindhoven Laboratory.

In projectors giving relatively small pictures (35 cm × 46 cm and 22 cm × 29 cm) use is made of experimental projection tubes having the same

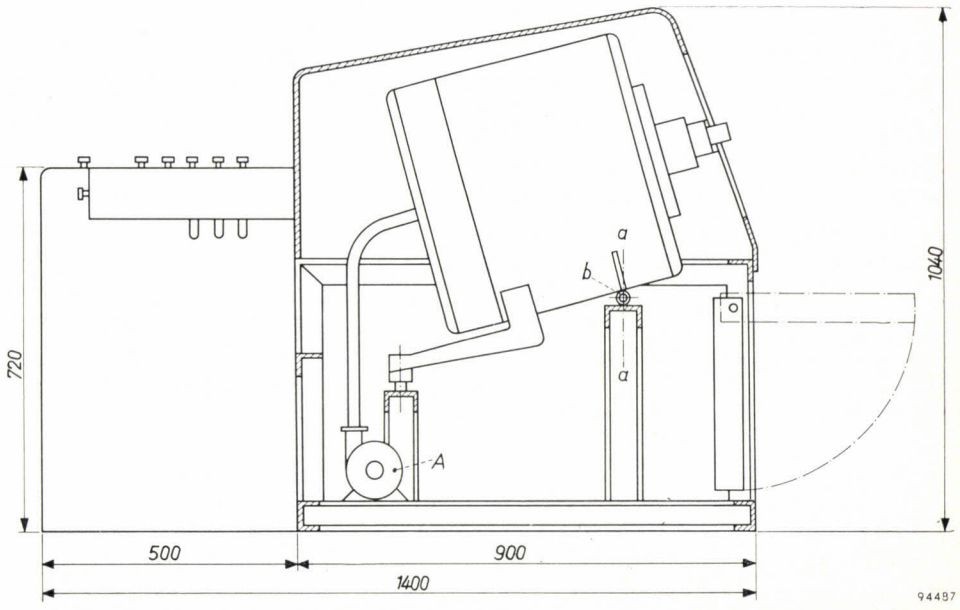


Fig. 21. Sketch showing the principal dimensions of the large-screen projector shown in figs. 20 and 22. Controls within the operator's reach allow him to turn the primary-colour projectors about a vertical axis *a-a* (in a ball and socket joint) and about a horizontal axis *b* (perpendicular to the plane of the drawing). *A* is the fan for cooling the projection tubes.

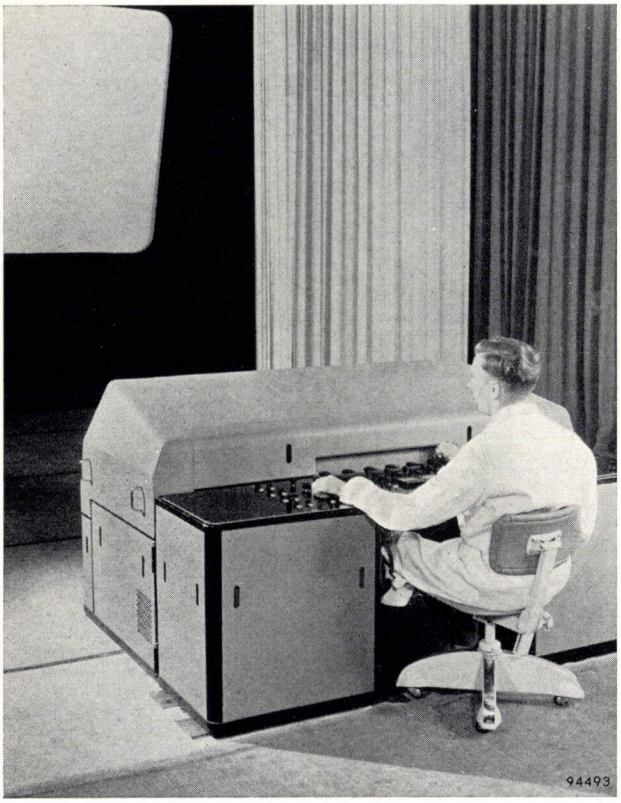


Fig. 22. The projector of fig. 20 in its housing. The picture size is 2.25 m × 3 m (≈7½′×10′) and the maximum luminance 20 cd/m².

dimensions and supplied with the same voltages as the MW 6-2 monochrome tube [9]) (screen diameter 6 cm, anode voltage 25 kV). Each tube is mounted in a Schmidt optical system with a mirror at 45°, a system of the type originally developed for the MW 6-2 [2]). The light passes through crossed dichroic mirrors, is reflected by a further plane mirror inclined at 45° and falls on to a transmission projection screen of polyvinyl chloride. Grooves on the front and back of the screen give it the desired directional characteristic. *Fig. 19* shows a projection apparatus of this kind, intended for studio monitoring purposes. The picture measures 22 cm × 29 cm and has a maximum luminance of about 200 cd/m².

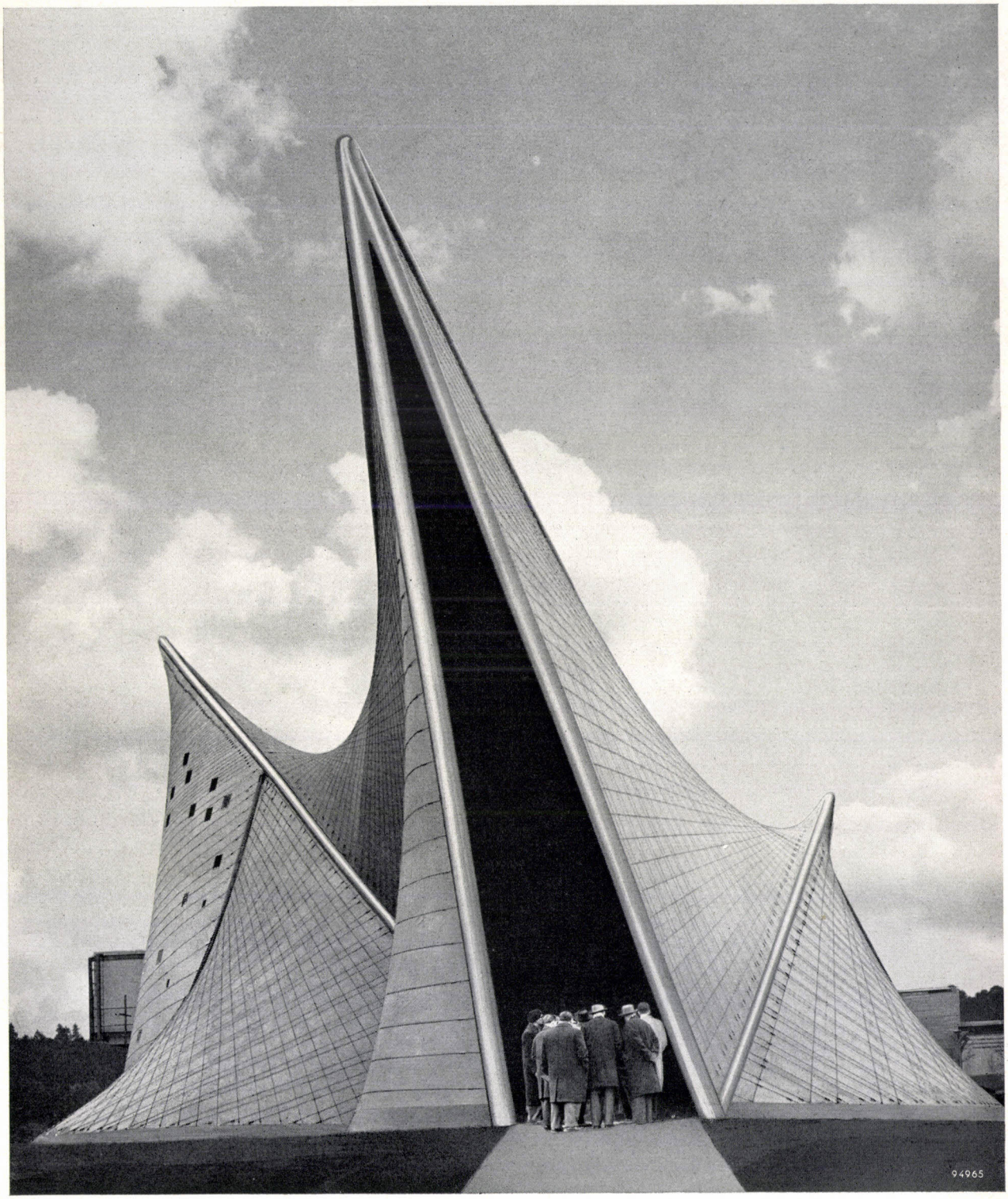
In equipment for large pictures (2.25 m × 3 m) use is made of experimental projection tubes with a screen of 13 cm diameter and an anode voltage of 50 kV. A photograph of a large-screen projection apparatus equipped with these tubes and with dichroic mirrors has been shown in fig. 4, earlier in this article. A maximum luminance of about 10 cd/m² is obtained on a beaded screen giving a "gain" of $\mu = 2.8$. Electrical adjustments are made by remote control, from a desk at the rear of the hall.

In another type of large-screen projector use is made of "in-line" Schmidt optical systems mounted side by side (*figs 20* and *21*). In this apparatus the control panel forms an integral part of the projector unit (*fig. 22*). Amongst the controls on the panel are six knobs for altering the alignment of the primary-colour projectors. These can be turned about two axes, a vertical axis *a-a* (fig. 21) and a horizontal axis through the points *b*. The apparatus embodies neither dichroic mirrors nor 45°-mirrors, giving as a result a maximum luminance of about 20 cd/m² on the screen (again a beaded screen with a μ of 2.8). Moreover, definition and contrast are considerably better in this projector than in that of fig. 4, and its height (fig. 21) is a good 30 cm less.

Summary: In the reproduction of colour television pictures by projection methods, three cathode-ray tubes of the projection type are used having respectively red, green and blue fluorescing phosphors. Each tube is mounted in a Schmidt optical system, either of the "in-line" type or of the "folded" type. The article is largely devoted to discussing two methods of obtaining superposition of the three primary-colour images. In the first method dichroic mirrors are employed in combination with folded optical systems. All three images are virtually projected from a single point; in theory, therefore, they are superimposed from the outset. This is not the case in the method where the primary-colour projectors are mounted side by side, but not parallel to one another; here the two outside tubes project images of trapezium shape. The trapezia can be converted into rectangles by making certain modifications to the scanning currents of the outer tubes. The current waveforms required for trapezium correction are discussed, as also the manner of generating them. Other matters dealt with in connection with scanning are skew correction, the adjustment of amplitude, and the displacement and rotational alignment of the images on the projection screen. Extra high tension stabilization methods, proportional brightness control and the characteristics of the video amplifiers are also discussed. Finally, some details are given of a number of colour television projectors actually constructed in the Philips laboratory at Eindhoven. Of two studio monitoring sets whose optical systems embody dichroic mirrors, one projects a picture measuring 22 cm × 29 cm, the other a picture measuring 35 cm × 46 cm; the maximum luminance is 200 cd/m². One large-screen projection set embodying dichroic mirrors gives a picture 2.25 m × 3 m with a maximum luminance of 10 cd/m²; another, in which the primary-colour projectors are mounted side by side, gives a picture 2.25 m × 3 m with a maximum luminance of 20 cd/m².

[9]) J. de Gier, Philips tech. Rev. **10**, 97-104, 1948/49.

THE PHILIPS PAVILION AT THE BRUSSELS WORLD EXHIBITION



In this building, designed for Philips by the architect Le Corbusier and his collaborator Y. Xenakis, an "Electronic Poem" is performed for visitors to the Brussels Exhibition. It is a composition of light and sound, the scenario being by Le Corbusier and the Music by Edgar Varèse. The unusual but effective architectural conception of the building and its construction in prestressed concrete will be the subject of a number of articles in one of the following issues of this Review. It is also the intention to publish a subsequent article on the technical aspects of the "Electronic Poem".

THE GENERATION OF DISLOCATIONS BY THERMAL STRESSES

by P. PENNING.

548.4:546.289

The term dislocation, denoting a particular kind of lattice imperfection in crystalline materials, was first introduced about 20 years ago. Since then dislocations have lost their hypothetical status — their existence is now well-established and they play an essential role in the behaviour of crystalline solids.

Dislocations (and other lattice imperfections) have a marked influence on the physical and mechanical properties of solids. This is of particular interest in the semiconducting materials germanium and silicon, which are nowadays of considerable technical importance. It is therefore not surprising that dislocations and their properties are the subject of a great deal of research. The present article deals with the formation of dislocations during the cooling of a germanium crystal, the study of which is important both with regard the problem of how to control the properties of germanium more effectively, and in order to throw fresh light on the origin of dislocations.

The preparation of single crystals of germanium

The importance of the elementary semiconductor materials germanium and silicon for various electronic devices is well known. Before these materials can be put to practical use, however, for example in the manufacture of transistors, they must be prepared in the form of single crystals of well-defined composition. This is generally done by causing the molten material to solidify under carefully controlled conditions. Of the various methods that have been developed we shall briefly describe here the Czochralski method, and confine ourselves mainly to the material germanium, though much of what will be said is also applicable to other methods and to silicon.

In the Czochralski method a seed crystal, mounted at any desired orientation, is lowered into a crucible containing molten germanium. If the temperature of the molten germanium is only slightly above the melting point, the seed, when pulled slowly upwards, will begin to grow and a single-crystal of cylindrical shape is formed. As the crystal is pulled out of the melt the solidified material cools off. Since in general the dimensions of the withdrawn crystal rod are fairly large (generally one or more cm thick) the inner material cools down more slowly than the material at the surface. This gives rise to temperature differences between axis and periphery, the magnitude of which are determined by the size of the crystal rod, the thermal conductivity of the crystal, the rate of withdrawal, the constitution of the gas ambient in which the growth takes place, etc.

The greatest temperature differences will generally occur immediately above the plane of solidification and may amount to as much as 40 °C. As a result of the thermal expansion, the central regions, being at a higher temperature will try to expand. This, however, is opposed by the surrounding material. The consequence is a fairly complex state of deformation, which can best be described by considering the stresses accompanying the process (forces between warmer and colder parts).

Such thermal stresses are of course present not only during the cooling of the material from the melt; they will always be present when a material is unevenly heated or cooled. Apart from the obvious case of a uniform temperature distribution, there is only one non-uniform temperature distribution which can be shown to give rise to no internal stresses, namely that at which the temperature *gradient* in the material is the same everywhere.

The thermal stresses are accompanied in the first place by an elastic, reversible deformation; as soon as the temperature again becomes uniform the deformation disappears. If the thermal stress is large, two subsidiary, permanent effects may appear. Firstly, if the material is brittle it may fracture. An example will be given below. Secondly, if the material is ductile it may be plastically deformed. This deformation is associated with the occurrence of lattice imperfections, such as dislocations, vacancies, etc. [1]), and hence destroys to some extent the perfection of the single crystal. How large the plastic deformation is, compared with the elastic deformation, depends entirely on the properties of the material and on the magnitude of the stress. Since the introduction of lattice imperfections generally has an adverse effect on the material, the aim during growth and all other treatments to

[1]) For a discussion of the mechanism of the plastic deformation of crystals, see, for example, H. G. van Bueren, Lattice imperfections and plastic deformation in metals, Philips tech. Rev. **15**, 246-257 and 286-295, 1953/54.

which the crystal is subjected is to keep the thermal stresses as small as possible. To avoid them entirely is not possible in practice.

As regards one kind of lattice imperfection viz. dislocations, it is fortunate that in a number of cases, including that of germanium, they can be made visible. When a polished germanium surface is etched with a special agent, almost every dislocation terminating at the surface gives rise to an etch pit (*fig. 1*). These etch pits can be counted and their distribution over the surface studied. This makes it possible not only to discover the concentration of these imperfections but also to investigate the conditions under which they arise, and hence to modify these conditions for the better.

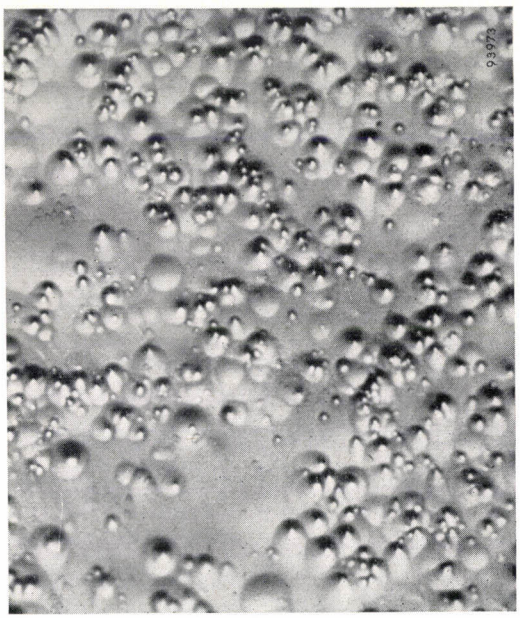


Fig. 1. Etch pits on the surface of a slightly deformed germanium crystal. The pits indicate the places where dislocation lines terminate at the surface (magnification 75×).

In the following we shall first consider the thermal stresses produced in germanium during certain experiments, and their effects on the etch-pit pattern. After a digression into the theory of the plastic deformation of germanium, we shall examine the information to be derived on the nature of this deformation from the observations described.

Thermal stresses in germanium rods during cooling

As will be clear from the foregoing, during the cooling of a hot cylindrical rod the cooler outermost parts of the rod will try to shrink more than the average and the warmer innermost parts less. As a result, the outermost parts will be subject to a tensile stress and the innermost parts to a compressive stress.

If the cylinder is cooled very rapidly (quenched), for example by plunging it into a bath of much lower temperature and good heat conductivity, stresses will *initially* be set up only at the surface, but these will be very large indeed. We can estimate these stresses as follows. Assume that, immediately after quenching, the mass of the material is still entirely at the initial high temperature T_0, but that the surface has taken on the quench temperature T_f. A very thin layer at the surface will then try to shrink by the relative amount $\varepsilon = \alpha(T_0 - T_f)$ where α is the coefficient of thermal expansion. It is prevented from doing so by the underlying material, and thermal stresses are set up (in this layer alone) which oppose the shrinkage. Since there are no external forces acting on the crystal, these can only be tangential (suffix t) and axial (suffix z) stresses in the surface layer.

From the theory of elasticity we know that these stresses amount to

$$\sigma_t = \sigma_z = \frac{E}{1-\nu}\alpha(T_0 - T_f), \qquad (1)$$

in which E is Young's modulus and ν is Poisson's ratio. According to this formula, immediately after the quenching of a germanium crystal over several hundreds of degrees, stresses will be set up in the surface layer amounting to some tens of kg/mm². These can easily exceed the critical stress for cleavage, which for germanium is about 20 kg/mm². Hence surface cracks will appear in the germanium rod, which accommodate the thermal stresses to such an extent that elsewhere in the rod no appreciablic plastic deformation occurs. This was confirmed by the following experiment. A germanium rod at a temperature of 500 °C was suddenly plunged into a water bath. Theoretically, according to (1), this would produce stresses of approximately 50 kg/mm² After the experiment the rod showed a large number of surface cracks (*fig. 2*); examination under the microscope revealed that, to judge from the number

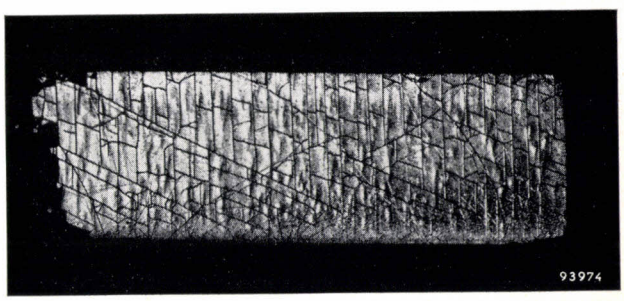


Fig. 2. Surface of a germanium rod quenched from 500 °C in a water bath. The crystal becomes covered with a large number of surface cracks (actual size).

of etch pits, no significant increase in dislocation density had taken place between the cracks.

Entirely different results are found if a germanium crystal is cooled relatively slowly, for example in a stream of gas. The whole process now takes place much less abruptly: the heat flux (which, in a long cylinder, is mainly radial) is much smaller, and in order to evaluate the maximum stress we must take time effects into account. Following the classical methods of solving heat-conduction problems, we can plot the deformations tending to arise from temperature differences as a function of time and place, provided the heat flux at the surface is known. The thermal stresses can then be ascertained with the aid of elasticity theory. One again finds [2]) that in the very first moments the stresses occur principally in the outermost surface layer, but fairly soon afterwards, before they have reached their maximum value there, stresses begin to arise also in the innermost parts. It appears that when the maximum value is reached at the surface the stresses in the centre are also fairly near maximum. This stress pattern builds up so quickly that the average temperature has dropped only very little by the time the stresses have reached their maximum value. From that point onwards the stresses everywhere start decreasing almost exponentially with time, but there is no further change in the *relative* magnitude of the stresses, i.e. the stress pattern then existing remains as it is, only the absolute values decreasing. The rate of this decrease may be expressed in terms of a characteristic time which is proportional to the radius of the rod, and to the specific heat per unit volume, and inversely proportional to the thermal conductivity of the material. This characteristic time varies between a few seconds and a few tens of seconds in the cases considered here.

The stress distribution set up after the initial period is parabolic across the diameter of the rod and is represented by the following equations (using polar coordinates r, Θ, z)

$$\left.\begin{aligned} \sigma_r &= \left(\frac{r^2}{R^2} - 1\right)\sigma_0, \\ \sigma_\Theta &= \left(3\frac{r^2}{R^2} - 1\right)\sigma_0, \\ \sigma_z &= \left(4\frac{r^2}{R^2} - 2\right)\sigma_0, \end{aligned}\right\} \quad \ldots\ldots \quad (2)$$

where R is the radius of the crystal rod and σ_0 is a term giving the exponential decrease of the stresses with time. The value of σ_0 is given by an expression of the form (1), modified by a factor β which depends on the cooling rate and the dimensions of the rod ($\beta \approx 0.01$ in the tests described below).

Now for the occurrence of plastic deformation — and hence the introduction of lattice imperfections — only the shear stresses acting on certain crystallographic systems are of importance (see[1]) and below, next heading) since a purely hydrostatic pressure (equal compression from all directions) does not give rise to plastic deformation. We may therefore, for convenience, subtract any hydrostatic pressure from equations (2) without affecting the problem. We choose to subtract a pressure equal to σ_Θ, since this enables us to dispose of the second of equations (2). We are left with only two normal stress components, σ_r in the radial direction and σ_z in the axial direction:

$$\sigma_r = -2\frac{r^2}{R^2}\sigma_0, \quad \ldots\ldots \quad (3)$$

$$\sigma_z = \left(\frac{r^2}{R^2} - 1\right)\sigma_0. \quad \ldots\ldots \quad (4)$$

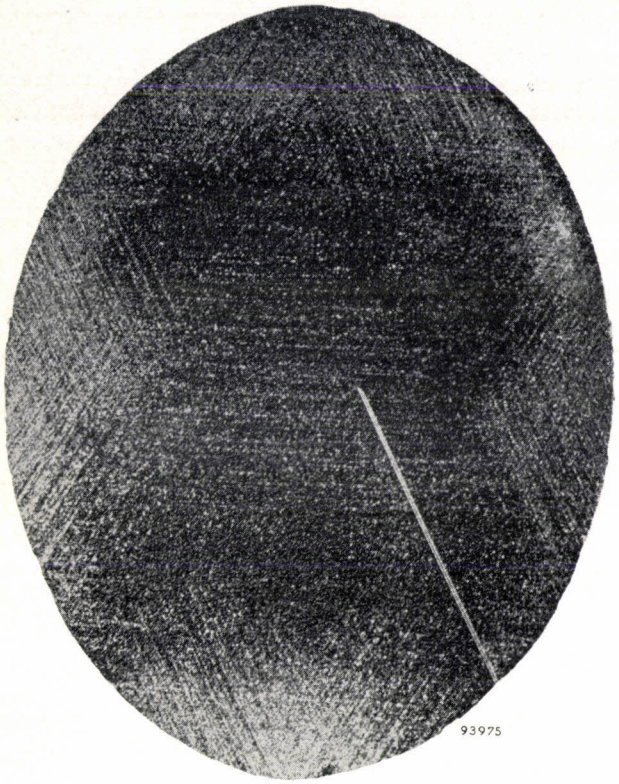
These stresses are both compressive: the radial compression is greatest at the surface and zero at the axis while the axial compression is maximum along the axis and zero at the surface. The maximum values of σ_r and σ_z are $2\sigma_0$ and σ_0 respectively. The value of the factor β being about 0.01, the maximum stress was smaller by approximately this factor than the value 50 kg/mm^2 calculated for rapid quenching, and amounted to a few kg/mm^2. This is of the order of the yield point of germanium at temperatures of about 500 °C. In these circumstances we may expect that at least a part of the thermal stress will be eliminated by plastic flow, and dislocations will be formed.

The following experiments demonstrated that this in fact took place. Germanium crystals of rod-shape, with their axes in a known crystallographic orientation, were heated to 850 °C and then cooled ("quenched") in a stream of hydrogen. Microscopic examination, after etching of a cross-section perpendicular or oblique to the axis of the rod, showed the presence of a typical regular pattern of etch pits, examples of which can be seen in *Fig. 3a* and *b*. The etch pits lie on rows whose direction is found to coincide with the intersections of (111) planes (octahedral planes) with the etched surface [3]).

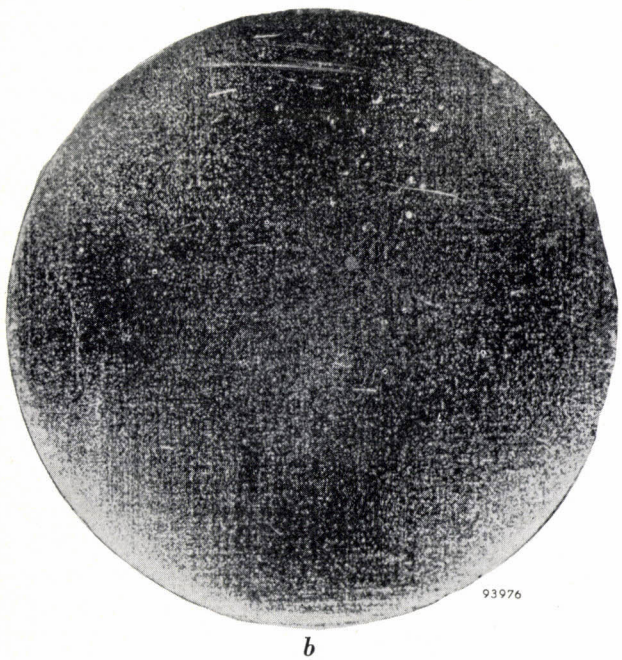
[2]) P. Penning, Generation of imperfections in germanium crystals by thermal strain, Philips Res. Repts. **13**, 79-97, 1958 (No. 1).

[3]) In this article no distinction is made between the symbols (*hkl*) and {*hkl*}, which denote respectively a single plane and the whole family of crystallographically equivalent planes. Nor is any distinction made between the symbols [*hkl*] and <*hkl*>, which refer in a similar way to crystallographic directions.

Which of the four possible (111) planes are involved at any particular region depends on the position of that region in the cross-section.



a



b

Fig. 3. *a*) Etch-pit pattern produced after etching on an oblique (111) cross-section of a cylindrical crystal with its axis in a [110] direction, which was cooled in a stream of gas. The etch pits lie in rows parallel to the intersections of the surface with the three other (111) planes. Note the minimum in the etch-pit density about midway along the radius of the cross-section. *b*) Etch-pit pattern for a perpendicular (100) cross-section of a [100] crystal, cooled as (*a*).

The average number of etch pits per cm^2 is hundreds of times larger than it was before heating to 850 °C and cooling; the number is largest at the periphery and smallest in a zone roughly half way along the radius of the cross-section.

These striking aspects of the distribution of the etch pits indicate that the dislocations do indeed arise during plastic deformation due to thermal stresses resulting from cooling.

In order to give a better foundation for this conclusion we shall devote a few words to the mechanism of the plastic deformation of germanium.

Plastic derformation of germanium

Germanium is extremely brittle at room temperature. Under sufficient stress it usually breaks as a result of cleavage along (111) planes, like diamond, to which it is structurally related. It has recently been found that at temperatures above about 400 °C germanium begins to exhibit a certain degree of ductility when the material is subjected to stresses of a few kg/mm^2. The higher the temperature at which the stress is applied the better is the germanium able to yield to this stress by plastic deformation until, near the melting point, it is very ductile. An interesting point is that, where the temperatures are not too high, a certain delay effect is observed: plastic deformation of the germanium occurs only some time after the application of the stress. For example at 400 °C, after application of stress, there may be an incubation period of several minutes before a noticeable deformation begins to take place, and even then it shows the typical character of creep, i.e. it increases only gradually with time. At higher temperatures the delay time decreases very rapidly, and above about 600 °C it is practically no longer perceptible, the deformation taking place almost immediately after the stress is applied.

Microscopic examination of a surface polished beforehand reveals that the mechanism of plastic deformation in germanium is principally a slip mechanism. Slip lines become visible, as shown in *fig. 4*. These slip lines are found to be oriented along intersections with (111) planes and the direction of slip always coincides with a [110] direction. This points very clearly to a dislocation mechanism as observed in metals [1]): slip in germanium evidently takes place owing to the formation and movement of dislocations with Burgers vectors along [110] directions over slip planes which have (111) orientations.

The germanium lattice has the same structure as that of diamond. *Fig. 5* shows the elementary cell of the diamond lattice. It will be clear that the (111)

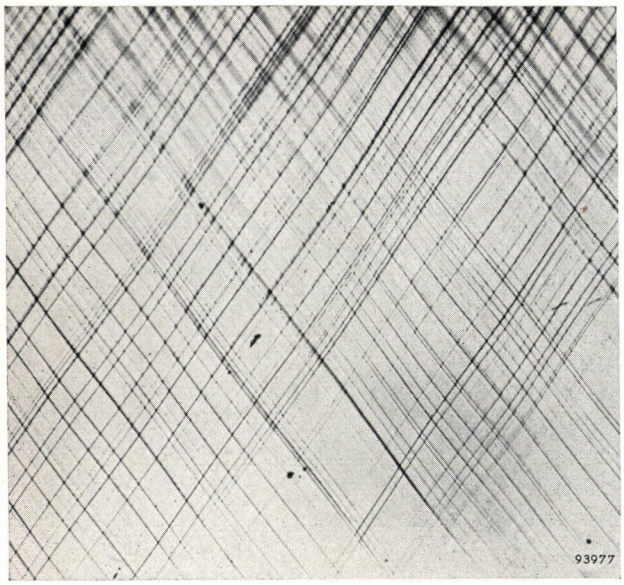


Fig. 4. Slip lines on a (110) plane of a plastically bent germanium crystal (75×). The lines run parallel to (111) planes.

planes are planes of closest packing, and moreover, that the shortest lattice period will be encountered in the [110] directions. On this basis one can understand theoretically the reason for the observed dislocation behavior: the dislocations occurring are those having the lowest energy.

It is to be expected that the density of the dislocations appearing at the surface is greatest in the slip lines. By suitably etching a surface covered with slip lines we will thus obtain rows of etch pits, again running along intersections of the (111) planes with the surface. This is illustrated in *fig. 6*. On the etched face, which is (100), two mutually perpendicular systems of rows of etch pits are visible, corresponding to two sets of (111) slip planes.

The formation and movement of dislocations on a particular slip system — i.e. combination of slip plane and slip direction (Burgers vector) — will occur only providing the shear stress component in the direction of the Burgers vector of this system exceeds a certain critical value. For a given state of stress, this condition will, generally speaking, be satisfied only on a few of the twelve possible slip systems in the diamond lattice. For this reason, the state of stress during the cooling of a cylindrical crystal will give rise to slip only on certain of the possible slip systems, depending on the crystallographic orientation of the axis of the rod. After etching, therefore, a quite specific etch-pit pattern may be expected. We shall now inquire into how far the etch-pit patterns observed after slow cooling are in agreement with these considerations.

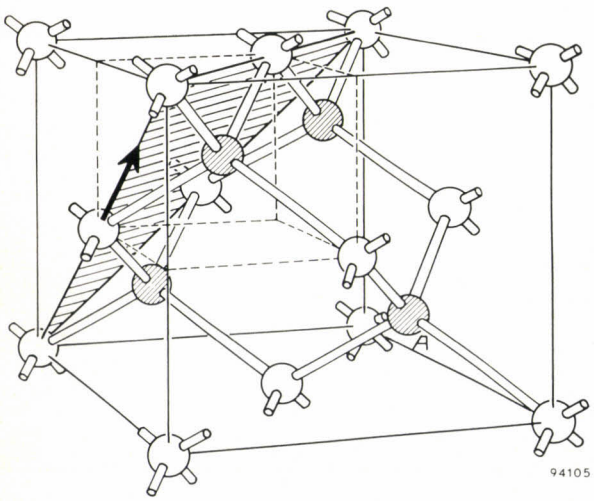


Fig. 5. Unit cell of the diamond structure, the crystalline structure of germanium. The hatching indicates one of the four (111) planes, and the arrow one of the six [110] directions.

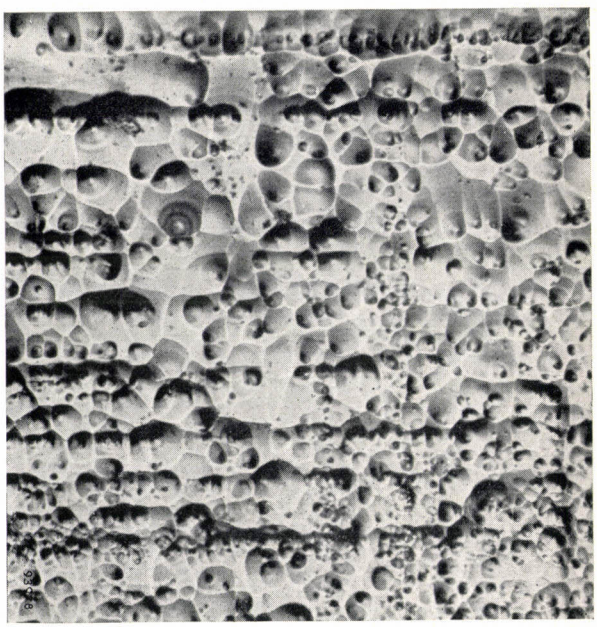


Fig. 6. Etch pits produced after etching a (100) plane, containing slip lines, of a plastically bent germanium crystal. The rows of pits lie along the intersections with (111) planes i.e. along [110] directions.

Distribution of dislocations after deformation by thermal stresses

We have seen that, if the deformations can be regarded as purely elastic, the stress distribution can be approximately given by (3) and (4), at the moment that the stresses are at maximum. Let us now assume that it is this elastic stress distribution which determines the degree of plastic flow. This implies the assumption that only a small proportion of the stresses is eliminated by plastic deformation, so small that the stress distribution is hardly affected thereby. This assumption is by no means self-evident. In view of the considerable plasticity

of germanium at high temperatures, one might expect at first sight that almost all differences in thermal expansion would be accomodated by plastic deformation, and that, after cooling, the material would be approximately free of stresses. In that case,

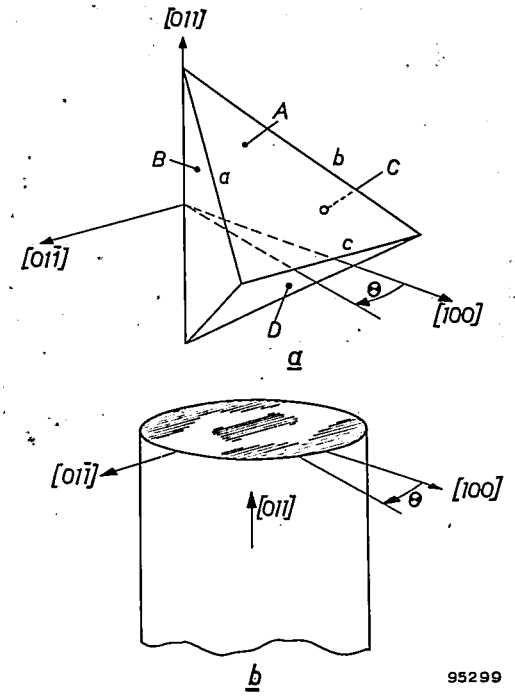


Fig. 7. *a*) The orientation of the four possible (111) slip planes. *A*, *B*, *C* and *D* with respect to the axis of a [011] crystal. The three slip directions in the plane *A* are indicated by *a*, *b* and *c*. *b*) [011] crystal rod cut parallel to (011) plane showing roughly the regions (hatched) where slip on planes *A* takes place (see further fig. 8).

of course, (2) would not apply, these formulae being derived from elasticity theory alone. An entirely different distribution of deformations and displacements would then appear instead of the deformations due to elastic thermal stresses, one reason being that plastic flow is only possible in certain directions. This distribution, too, can be calculated. However, for reasons stated below, we shall first confine ourselves to the distribution due to purely elastic stresses and consider a rod-shaped single crystal with a [110] direction parallel to the axis. Fig. 7 shows schematically the orientation with respect to this axis of the four (111) planes, each of which has three [110] directions — giving the twelve possible slip systems. To define the position of any point in a cross-section perpendicular to the axis, two coordinates will be used, namely the distance r from the centre, and the angle Θ between the direction r and the [100] direction. We can now roughly ascertain in the following way when and on what planes slip will occur.

Near the periphery of the rod the radial normal component of stress, σ_r, is predominant. Only where this stress-component makes an angle not too near to 0° or 90°, both with the slip plane and the slip direction, does an appreciable shear stress component act in the slip system concerned, in which case slip may occur. (When a body is acted on by a uniaxial stress σ making an angle φ with the normal to the slip plane and an angle λ with the slip direction, the shear stress component in the slip direction is given by $\tau = \sigma \cos \varphi \cos \lambda$.) Consider the slip plane A (*fig. 7a*). This lies favourably for those positions in the cross-section where $\Theta \approx 0°$; here only the slip directions a and b on plane A can be active. For positions corresponding to $\Theta \approx 45°$, the slip direction c becomes favourably oriented. For $45° < \Theta < 90°$ the shear-stress component resolved in the plane A becomes increasingly smaller, and slip becomes increasingly less probable, until at $\Theta = 90°$ the shear stress acting in this plane has become zero. In the other quadrants the variation with Θ is analogous. In the inner regions, near the axis of the rod, the axial stress σ_z predominates. The probability of slip on plane A is there practically independent of Θ. Since, however, the resolved shear stresses due to σ_r and σ_z have opposite sign, the net deformation is zero in those regions where both stresses are of approximately equal magnitude, i.e. about midway along the radius: thus no slip occurs on plane A in an annular region (fig. 7*b*). The annular slip-free region is not independent of Θ. At $\Theta = 90°$ and 270° the radial stress component vanishes, as we have seen, so that the outer boundary

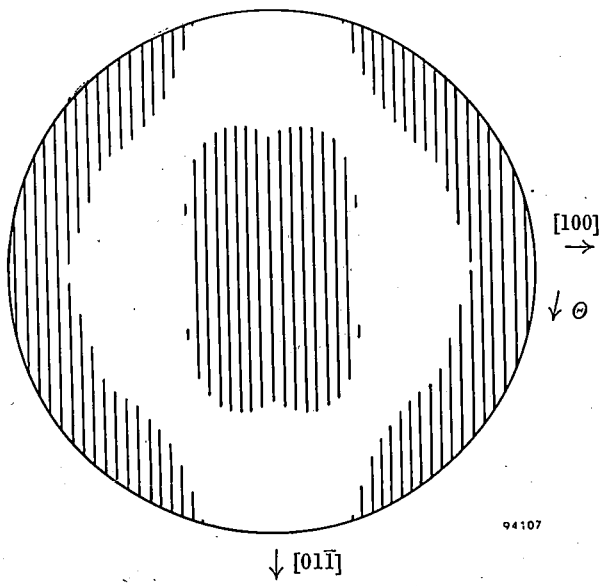


Fig. 8. Cross-section of a[110] crystal showing the regions in which slip on the slip plane A (fig. 7) is to be expected due to thermal stresses (i.e. regions where the shear stress in at least one slip direction exceeds a certain value). Between a roughly oval slip zone in the centre and a peripheral zone consisting of two parts, lies a region where there is practically no slip at all.

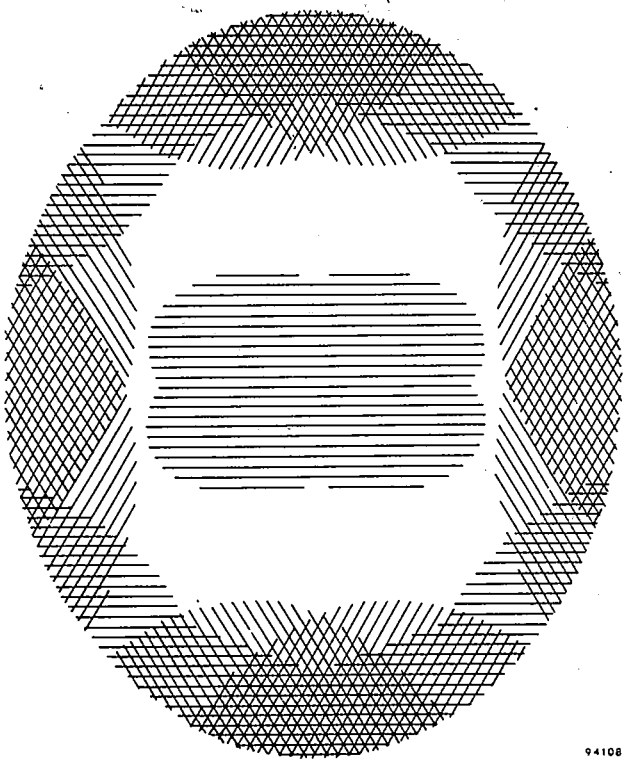
of the slip-free region extends to the edge of the crystal; also the resolved stress due to the axial component gives rise to slip at points further from the axis in the directions 90° and 270° where the opposing stress due to σ_r is smaller, so that the inner boundary of the slip-free region is roughly oval in shape.

In this way we can indicate roughly in what parts of the cross-section the resolved shear stress will exceed a certain value along at least one of the slip directions of plane A (say), i.e. in what zones slip will take place on plane A, and hence give rise to etch pits at the surface. It should be mentioned that the value chosen for this resolved shear stress is not critical, that is to say, slip patterns of similar form are obtained for a range of values of this stress. Accordingly a value is chosen which leads to slip patterns which conform in *size* with those observed. The results, as regards slip on plane A, are given in *fig. 8*. The direction of the hatching indicates the orientation of the effective slip plane A. As regards plane D, of course, the same figure applies. The slip planes B and C give rise to other patterns which are displaced about 70° (the tetrahedral angle) with respect to each other, but are otherwise similar to each other; in this case only the radial stress component is of influence.

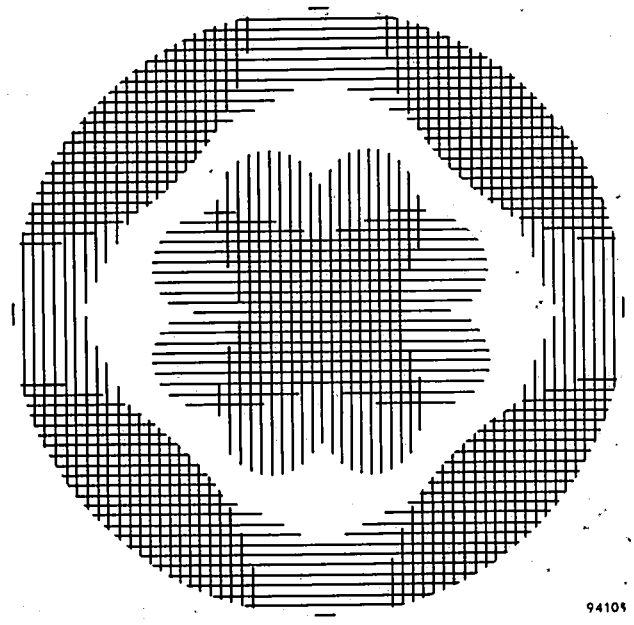
In order to be able to make some predictions regarding the pattern of the etch pits in a given cross-section we must superimpose the individual distributions obtained in this way, taking into account that only those dislocations will give rise to etch pits that lie in planes which do not make too small an angle with the plane of the cross-section. For an elliptical (111) cross-section (// to D, say) of a [110] rod only the dislocations associated with slip in the planes A, B and C will be observed. The result of the above considerations is represented in *fig. 9a* for such a (111) cross-section. This theoretical figure can now be compared with the observed pattern shown in fig. 3*a*. They agree very satisfactorily, both as regards the radial distribution of the etch-pit density (minimum in annular region roughly half-way along the radius, owing to the substantial equality of the opposing resolved components of axial and radial stresses), and as regards the azimuthal variations at the periphery (minima at an angle of about 45° with the axes of the ellipse).

Theoretical etch-pit figures for the other orientations investigated can be derived in an analogous way. An example is shown in fig. 9*b*, which should be compared with the experimental pattern in fig. 3*b*. Again, the agreement is very satisfactory.

These results thus indicate that we were justified in making the assumption that only a slight degree of plastic deformation occurs during cooling. The alternative, namely that complete plastic stress-relaxation occurs, appears on the other hand to lead



a



b

Fig. 9. *a*) Superposition of the theoretical slip patterns (such as that in fig. 8) for three slip planes of a [110] crystal. Since the (elliptical) cross-section is along a (111) plane, dislocations associated with slip in the fourth slip plane are not observable on this surface. The direction of the hatching indicates the orientation of the relevant slip planes. The pattern of this figure should be compared with the observed pattern of rows of etch pits in fig. 3*a*.
b) Theoretical slip pattern for a (100) cross-section of a [100] crystal. Compare with fig. 3*b*.

to theoretical etch-pit distributions (*fig. 10*) which differ considerably from those actually observed. (It may be noted here that a number assumptions concerning the plastic deformation of germanium

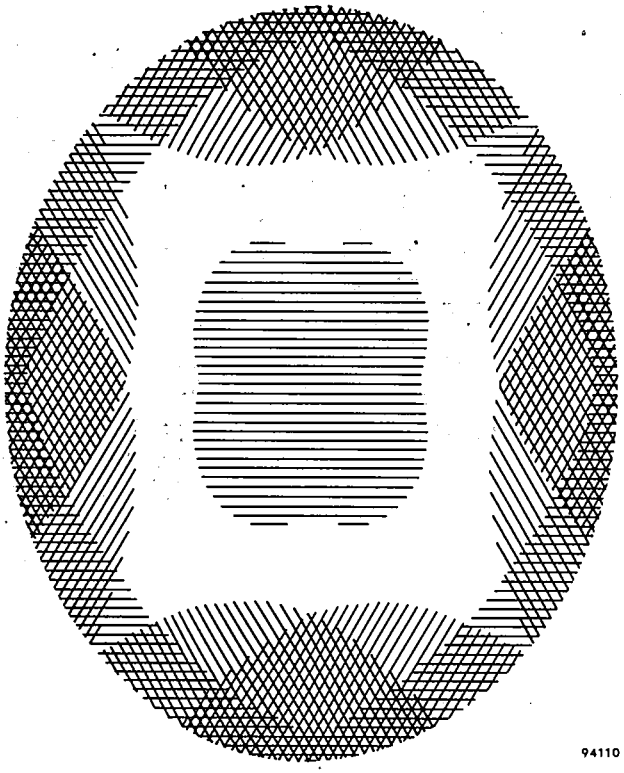


Fig. 10. Theoretical "slip figure" for the orientation in fig. 3*a* and fig. 9*a*, derived on the assumption of complete stress relaxation by plastic flow. The agreement between this figure and fig. 3*a*, is not nearly as good as between fig. 9*a* and fig. 3*a*, as appears, for example, from the shape of the central maximum and the position of the peripheral minima (the latter are almost absent in the figure above).

are necessarily involved in the derivation of fig. 10. The assumptions seem reasonable but their validity has not actually been verified.)

It may be asked why much greater plastic deformations do not occur. A possible answer might be based on the delay in plastic flow already mentioned, which characterizes the formation of new dislocations in germanium under the influence of deforming stresses. Although this delay time can no longer be measured in normal deformation tests above 600 °C, it can be shown that in the very short times elapsing during the first stages of cooling of a germanium rod (e.g. 0.1 sec. for the first 100 °C cooling) the number of dislocations that can be formed under the influence of thermal stresses does not amount to more than about 10% of the minimum number that would be necessary for a complete stress-relaxation. The cause of this delay in plastic flow is probably to be sought in the homopolar bonding forces between the atoms in the germanium lattice, which obstruct the movement of dislocations, or in an analogous influence arising from impurities, or in both.

Thermal stresses during crystal growth

From the good agreement between the observed etch-figures on quenched rods and the theoretical patterns derived above from considerations of slip, we may conclude that dislocations will also arise during the withdrawal of crystals from the melt. The perfection of the crystals will be determined by the thermal stresses present, and can thus be influenced by modifying the conditions during growth. As mentioned early in this article, a unifrom temperature gradient in the crystal will entail no thermal stresses. This state, however, can never be achieved because losses due to external radiation and, of course, the movement of the crystal, give rise to a heat flux resulting in a non-uniform temperature gradient in the crystal. The simple theory treated in the foregoing is, of course, unable to give a quantitative description of the very complicated situation during the growth of the crystal. We might, in addition, introduce an axial temperature gradient into the theory but even then we can only partially approach the actual situation. Qualitatively, however, the theory gives a good explanation of several aspects of the formation of dislocations during the withdrawal process: for example, the effect on the dislocation distribution when the crystal is rotated — a method employed for producing a crystal of good cylindrical shape.

Summary. During the pulling of germanium and silicon crystals from the melt, a process widely employed in practice, a non-uniform temperature gradient arises. As a result, thermal stresses are formed in the material which may cause plastic flow or even fracture of the material Some experiments are described concerning the influence of the rate of cooling of a germanium rod on the state of internal perfection of the material. The latter can be judged, after polishing and etching, by the number of etch pits observed under the microscope and their distribution over a cross-section of the rod. These etch pits are formed at those places where a dislocation line terminates at the face of the cross-section. The distribution and density of the dislocations depend on the degree of plastic deformation, and this in turn depends on the spatial distribution of the thermal stresses. These stresses are derived for a cylindrical rod with radial heat flux. For various crystallographic orientations of the rod, theoretical patterns of etch-pit distribution are derived which are in very good agreement with the figures observed. The theoretical figures are based on the assumption that the stresses are only partially relieved by plastic flow. This assumption, evidently justified by the good agreement found, throws an interesting light on the properties of dislocations in homopolar substances such as germanium.