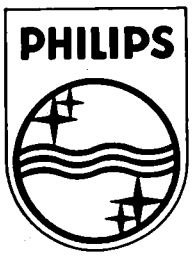
1668-68

PHILIPS TECHNICAL REVIEW



VOLUME 21

1959/60

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
D. van Dalen
K. R. Fröhner (German edition)
C. M. Hargreaves (English edition)
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 21, 1959/60

LIGHT

Motion-picture projection with a pulsed light source, by P. Hoekstra and C. Meyer . . . 73

A motion-picture projector of simplified design, by J. J. Kotte 83

Stray capacitances in neon installations, by J. J. Wilting 207

A simple circuit for a light source of constant intensity, by H. van Suchtelen 229

Temperatures in fittings for incandescent lamps, by L. J. H. Exalto 300

Automatic control of a filament-coiling machine with the aid of preset counters, by F. Einramhof and P. Havas 309

TELECOMMUNICATIONS

A range of pulsed magnetrons for centimetre and millimetre waves, by J. Verweel and G. H. Plantinga 1

A wide-band triode amplifier with an output of 10 W at 4000 Mc/s, by J. P. M. Gieles and G. Andrieux 41

Experiments in the field of parametric amplification, by B. Bollée and G. de Vries . . . 47

A slotted lecher line for impedance measurements in the metric and decimetric wave bands, by G. Schiefer 88

An 8 mm high-resolution radar installation, by J. M. G. Seppen and J. Verstraten . . . 92

A metal-ceramic disc-seal triode for frequencies up to 6000 Mc/s, by E. Mentzel and H. Stietzel 104

An experimental disc-seal triode for 6000 Mc/s, by M. T. Vlaardingerbroek 167

Reflex klystrons for wavelengths of 4 and 2.5 mm, by B. B. van Iperen 221

Flying-spot scanners for colour television, by H. van Ginkel 234

SOUND

Modern acoustical engineering, II. Electro-acoustical installations in large theatres, by D. Kleis. 52

An acoustic spectrum analyser with electronic scanning, by D. J. H. Admiraal 349

Performance tests on loudspeakers, by M. T. Haitjema, W. Kopinga and S. J. Porte . . 362

X-RAYS AND NUCLEAR PHYSICS

Instrumentation for a subcritical homogeneous suspension reactor,

I. Reasons behind the choice of a homogeneous suspension reactor, by J. J. Went. 109

II. Measurement and control of operating parameters, by B. L. A. van der Schee and M. van Tol 121

III A. The monitoring of low neutron flux by means of fast pulse-counting channels, by J. J. van Zolingen . . . 134

III B. The monitoring of high neutron flux with the aid of an electrometer, by M. van Tol 144

IV. The safety circuits, by F. J. Schijff . 148

The orientation of diamonds for tools by means of an X-ray image intensifier, by J. F. H. Custers and A. J. van der Wagt . . 178

A transistorized radiation monitor, by M. van Tol and F. Bregman 201

X-ray inspection of hot steel billets during rolling, by W. J. Oosterkamp, J. Proper and M. C. Teves 281

MEASURING EQUIPMENT

The resistance network, a simple and accurate aid to the solution of potential problems, by J. C. Francken 10

A slotted lecher line for impedance measurements in the metric and decimetric wave bands, by G. Schiefer 88

Instrumentation for a subcritical homogeneous suspension reactor,

II. Measurement and control of operating parameters, by B. L. A. van der Schee and M. van Tol 121

III A. The monitoring of low neutron flux by means of fast pulse-counting channels, by J. J. van Zolingen . . . 134

III B. The monitoring of high neutron flux with the aid of an electrometer, by M. van Tol 144

An automatic dew-point hygrometer using Peltier cooling, by P. Gerthsen, J. A. A. Gilsing and M. van Tol 196

A transistorized radiation monitor, by M. van Tol and F. Bregman 201

A gas-discharge indicator tube for transistorized decade counting circuits, by T. P. J. Botden 267

A transistor cardiotachometer for continuous measurements on working persons, by G. A. Harten and A. K. Koroncai 304

A simple method of determining the thermal conductivity of solids, by J. Schröder . . . 357

Performance tests on loudspeakers, by M. T. Haitjema, W. Kopinga and S. J. Porte . . 362

MATERIALS AND THEIR RESEARCH

Segregation and distribution of impurities in the preparation of germanium and silicon, by J. Goorissen 185

X-ray inspection of hot steel billets during rolling, by W. J. Oosterkamp, J. roPper and M. C. Teves 281

Nuclear magnetic resonance, by D. J. Kroon . 286

The fruits and foundations of solid-state research, by D. Polder 334

A method of growing dislocation-free germanium crystals, by B. Okkerse 340

A simple method of determining the thermal conductivity of solids, by J. Schröder . . . 357

MISCELLANEOUS

Vector-electrocardiography, by G. C. E. Burger and G. Klein 24

The use of radioactive isotopes for the study of littoral drift, by J. J. Arlman, J. N. Svašek and B. Verkerk 157

Metal vacuum equipment, by N. Warmoltz and E. Bouwmeester 173

The orientation of diamonds for tools by means of an X-ray image intensifier, by J. F. H. Custers and A. J. van der Wagt . . . 178

An automatic particle counter and sizer, by H. A. Dell, D. S. Hobbs and M. S. Richards 253

A gas-discharge indicator tube for transistorized decade counting circuits, by T. P. J. Botden 267

Experiments with radioactive preparations of the acaricide "Tedion V 18", by J. Halberstadt 276

Nuclear magnetic resonance, by D. J. Kroon 286

Automatic control of a filament-coiling machine with the aid of preset counters, by F. Einramhof and P. Havas 309

A 75 cm receiver for radio astronomy and some observational results, by C. L. Seeger, F. L. H. M. Stumpers and N. van Hurck . . 317

VOLUME 21, 1959/60, No. 1 pp. 1-40 Published 11th December 1959

Philips Technical Review

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

A RANGE OF PULSED MAGNETRONS FOR CENTIMETRE AND MILLIMETRE WAVES

by J. VERWEEL and G. H. PLANTINGA. 621.385.64.029.64:621.3.032.2

During the war years great strides were made in the development of the magnetron. Power outputs were raised from a few score watts to hundreds of kilowatts, and operating wavelengths reduced from some 20 cm to 1 cm. Progress since then has perhaps been less spectacular, but the following article, which describes a range of magnetrons for wavelengths of 32, 12, 8 and 4 mm, shows that advances are still being made in this field.

Introduction

In recent times the development of the magnetron has been mainly directed along the following lines [1]):

1) The achievement of higher pulse powers.
2) The attainment of shorter operating wavelengths.
3) The incorporation of tuning, i.e. increase of the frequency range.
4) The attainment of high continuous power outputs with high efficiencies.

Points 1) and 2) chiefly concern magnetrons for radar. With a shorter wavelength and the same size of aerial the received fraction of the transmitted power is greater and the beam narrower. This makes for better resolution of the image on the radar screen and allows the shape of objects to be better distinguished, which is important for short-range radar as used on airfields and in harbours. A radar installation of this kind equipped with an 8 mm magnetron will be the subject of a forthcoming article in this Review.

Point 4) may be of importance in shortwave therapy and also in electronic cookers, in which food is cooked or heated in a very short time by high-frequency radiation, usually of about 12 cm wavelength. Both applications are based on the high dielectric losses occurring in water at these wavelengths.

In this article attention will be devoted solely to points 1) and 2). We shall describe a range of experimental magnetrons designed at the Philips Research Laboratories, Eindhoven, for operation at wavelengths of about 32, 12, 8 and 4 mm. The peak power output of these tubes is, in round figures, respectively 1100, 70, 80 and 40 kW. In the case of the 32 mm magnetron the emphasis is placed on obtaining a high mean power as well as a high peak power. The other magnetrons may be regarded as steps on the way towards the highest possible operating frequency. All tubes are equipped with an L-type (dispenser) cathode, which has been found to give good results in magnetrons. The four tubes in question, together with some of their component parts, are shown in *fig. 1*, from left to right in order of diminishing wavelength.

Fig. 2 shows, from left to right, the 12 mm, 8 mm and 4 mm magnetrons each mounted in its permanent magnet. (No special magnet was developed for the 32 mm magnetron.)

In the design of magnetrons for ever higher frequencies, use can be made of the laws of similarity. This design technique will be discussed below. We shall then deal with the construction of the various tubes and with the engineering problems involved, and finally we shall compare the operating characteristics one with the other and in relation to the scaling laws derived from the similarity considerations.

[1]) For an introduction to the principles of magnetrons, see e.g. J. Verweel, Philips tech. Rev. **14**, 44-58, 1952/53. For a comprehensive treatment see G. B. Collins, Microwave magnetrons, Radiation Laboratory Series No. 6, McGraw-Hill, New York 1948.

Similarity considerations applied to magnetron design

The calculation of the parameters of a magnetron intended to meet specified requirements is in general so complex as to be virtually impracticable. For given dimensions and wavelength the theory (see [1])) postulates a minimum condition for the magnetic field B and also an approximately linear relation between B and the anode voltage V_a. Design follows primarily from experience and from analysis of existing magnetrons. The measured characteristics of a series of experimental magnetrons can then be used for making modifications until an optimum definitive design is evolved.

Another and more direct means of arriving at a new design, especially where it is the wavelength that is to be changed, is based on considerations of similarity. Suppose that a given magnetron at a given anode current I_a and anode voltage V_a requires a magnetic field B at which it delivers a certain power output with a wavelength λ. If we now scale up an existing magnetron so that its linear dimensions all become p times larger, then at the same current and voltage this magnetron will produce the same power output at the wavelength $p\lambda$, and for this it will require a magnetic field of strength B/p. If the similarity is exact, the specific conductance at a given position in the new magnetron should be $1/p$ times the specific conductance at the corresponding position in the old magnetron, and the same should apply to the electric field strength at the cathode.

The scaling laws appropriate to the various parameters of a magnetron can be deduced as follows. The charges, currents and fields in a magnetron can be found theoretically with the aid of Maxwell's equations (for the vacuum and the conductors) and the equations of motion for an electron:

$$\left.\begin{aligned} \operatorname{curl}\frac{\mathbf{B}}{\mu_0} &= \mathbf{S} + \varepsilon_0\frac{\partial \mathbf{E}}{\partial t} \\ \operatorname{curl}\mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\ \operatorname{div}\mathbf{E} &= \varrho/\varepsilon_0 \\ \operatorname{div}\mathbf{B} &= 0 \\ m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} &= e\mathbf{E} + e[\mathbf{v}\times\mathbf{B}] \\ \mathbf{S} &= \sigma\mathbf{E} \text{ in the conductors} \\ \mathbf{S} &= \varrho\mathbf{v} \text{ in the vacuum,} \end{aligned}\right\} \quad \ldots\ldots (1)$$

where $\mathbf{E}$ is the electric field strength, $\mathbf{B}$ the magnetic induction, $\mathbf{S}$ the current density, ϱ the charge density, $\mathbf{v}$ the velocity of an electron (charge e, mass m) and σ the electrode conductivity ($\mu_0 = 4\pi\times10^{-7}$ H/m and $\varepsilon_0 = 10^{-9}/36\pi$ F/m). The boundary conditions are given by the geometry and the conductance of the electrodes, the anode voltage and anode current, the constant axial magnetic field and the electric field at the cathode.

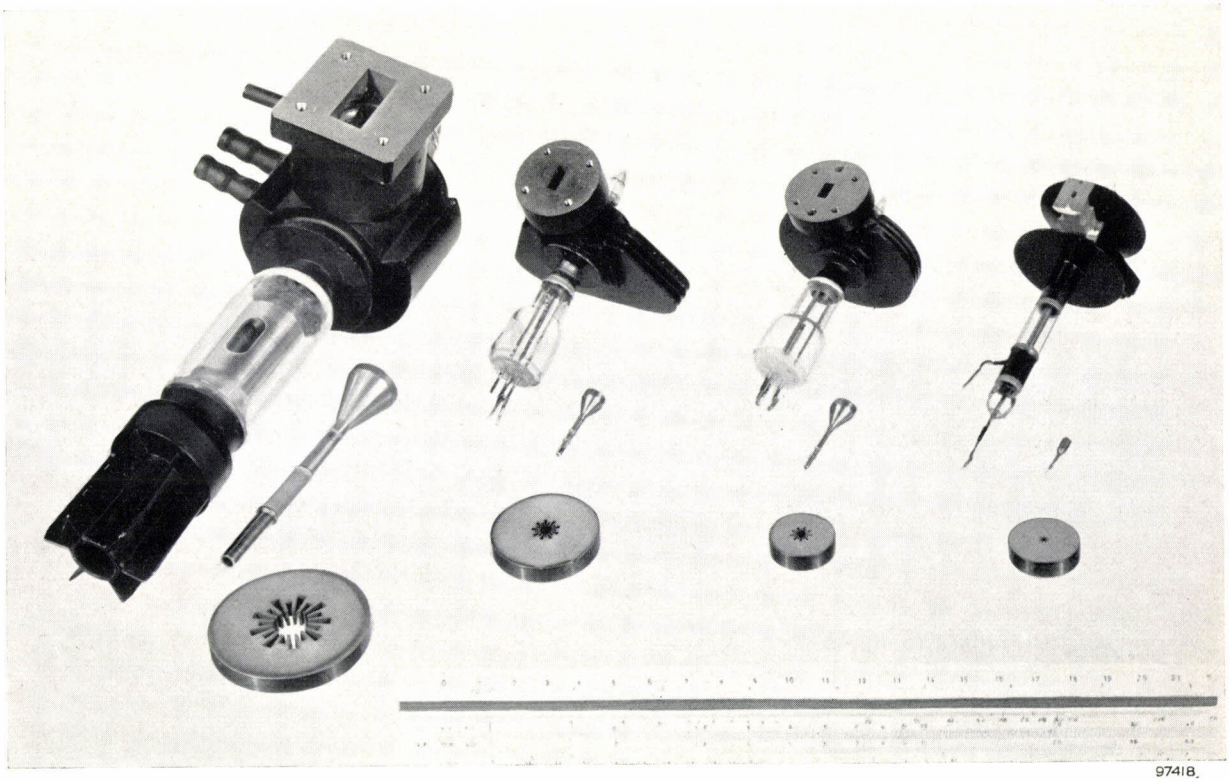
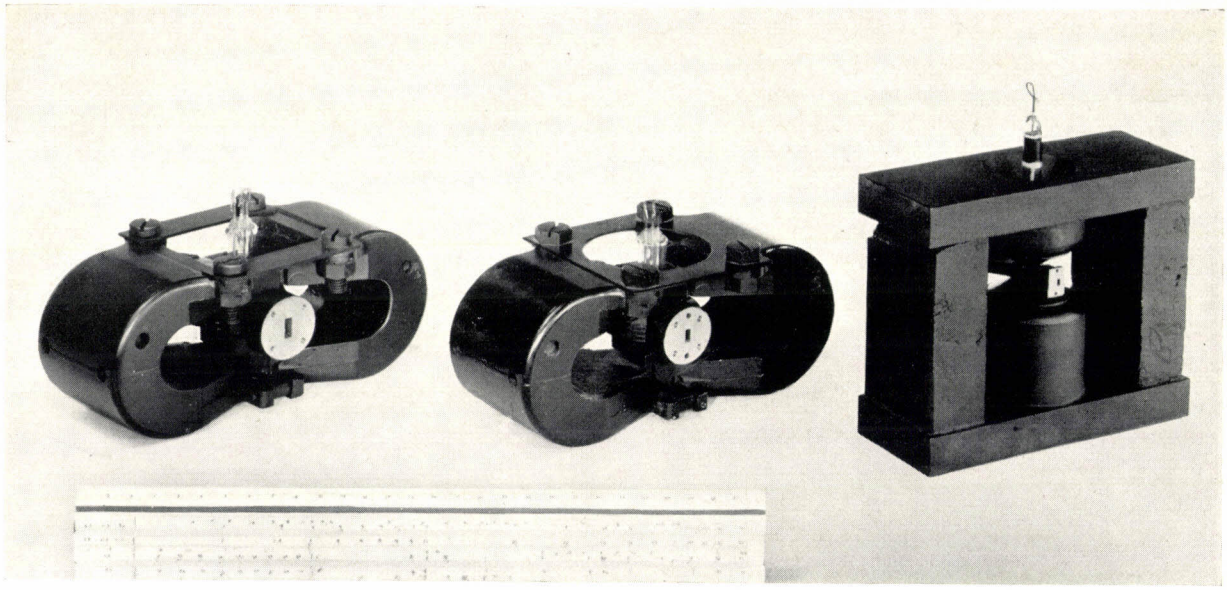


Fig. 1. The four magnetrons discussed here for wavelengths of (from left to right) about 32, 12, 8 and 4 mm. Also depicted are the copper anode blocks (not yet machined to the correct diameter and height), and the cathode assemblies.



97419

Fig. 2. Left to right: the 12, 8 and 4 mm magnetrons mounted in their permanent magnets.

We consider two magnetrons *1* and *2*; all the dimensions of *2* are larger than those of *1* by a factor p. Taking a coordinate system (x_1, y_1, z_1) for *1* and a coordinate system (x_2, y_2, z_2) for *2*, both with their origin in the centre of the magnetron, we can define corresponding points by:

$$x_2 = px_1, \qquad y_2 = py_1, \qquad z_2 = pz_1 .$$

We define moreover as corresponding times:

$$t_2 = pt_1 .$$

Let all electromagnetic quantities of *1* be known; we give these the suffix 1. If we give these quantities in magnetron *2* the suffix 2, and we write, for corresponding positions,

$$\mathbf{E}_2 = \frac{\mathbf{E}_1}{p}; \qquad \mathbf{B}_2 = \frac{\mathbf{B}_1}{p}; \qquad \mathbf{S}_2 = \frac{\mathbf{S}_1}{p^2};$$

$$\varrho_2 = \frac{\varrho_1}{p^2}; \qquad \mathbf{v}_2 = \mathbf{v}_1; \qquad \sigma_2 = \frac{\sigma_1}{p},$$

all these quantities then satisfy equations (1) for the magnetron *2*. This is indeed evident from the fact that the operators curl and div are combinations of first derivatives with respect to the coordinates. For example, the component

$$\frac{\partial \mathbf{B}_2}{\partial x_2} = \frac{\partial (\mathbf{B}_1/p)}{\partial p x_1} = \frac{1}{p^2}\frac{\partial \mathbf{B}_1}{\partial x_1}, \quad \text{etc.}$$

It is found that the potentials, the currents and the Poynting vector are the same in both magnetrons.

For the given boundary conditions the quantities with suffix 2 thus constitute the solution for magnetron *2*, the power output being the same. Since corresponding times in magnetron *2*, as defined above, are p times longer than in magnetron *1*, the frequency of the oscillations generated by *2* is $1/p$ times that of *1*.

The various scaling laws should thus enable us, if we have for example an existing 3 cm magnetron for the required power output, to construct a 6 mm magnetron by reducing all dimensions five times and by increasing the magnetic field five-fold. The application of similarity principles, however, can encounter difficulties, such as the following.

1) Towards higher frequencies it is not, in general, possible to increase the electrode conductance correspondingly, for with increasing frequency the effective conductivity in fact decreases as a result of skin effect. The question then arises in how far the real situation can be described by attributing an infinitely high conductance to the anode block of both magnetrons — in which case the scaling law would again be satisfied. As regards the effect on the electric field pattern, which determines the electron motion, the resistance of the conductors can normally be disregarded. Obviously, however, this is not so as regards the energy losses due to the high-frequency currents in the tube walls: in the smaller magnetron a larger proportion of the high-frequency power supplied by the electron beam will be needed to maintain the oscillations in the tube. We shall return to this matter when comparing the results.

2) The scaling factor for the magnetic induction is that it should increase proportionately with the frequency. This causes no difficulty while the magnetic induction is substantially lower than the saturation induction of magnetic materials, since the magnetic air-gap is also reduced. The length of the required magnetic material then remains approximately the same. However, if the induction required approaches the saturation value, prohibitive difficulties arise in the design of the magnet.

3) Since, under conditions of similarity, the current remains unchanged, the current density must increase as the square of the frequency. It may then happen that the cathode emission in the original

magnetron is limited by the space charge, whereas in the scaled-down magnetron the emission is saturated (and possibly supplemented by secondary emission to bring it to the required value; this will be touched on presently). The electric fields at the cathode may not be related in the two cases exactly as given by the relevant scaling factor, i.e. strict similarity no longer exists. With further reductions in size a limiting cathode current will be reached which will call for modifications to the design.

4) Upon scaling-down, the voltage remains unchanged, which means that the electric field strength increases. It is therefore necessary to take into account the danger of electrical breakdown.

5) The requirements with respect to mechanical strength and dimensional tolerances can lead to difficulties in the construction of a scaled-down model.

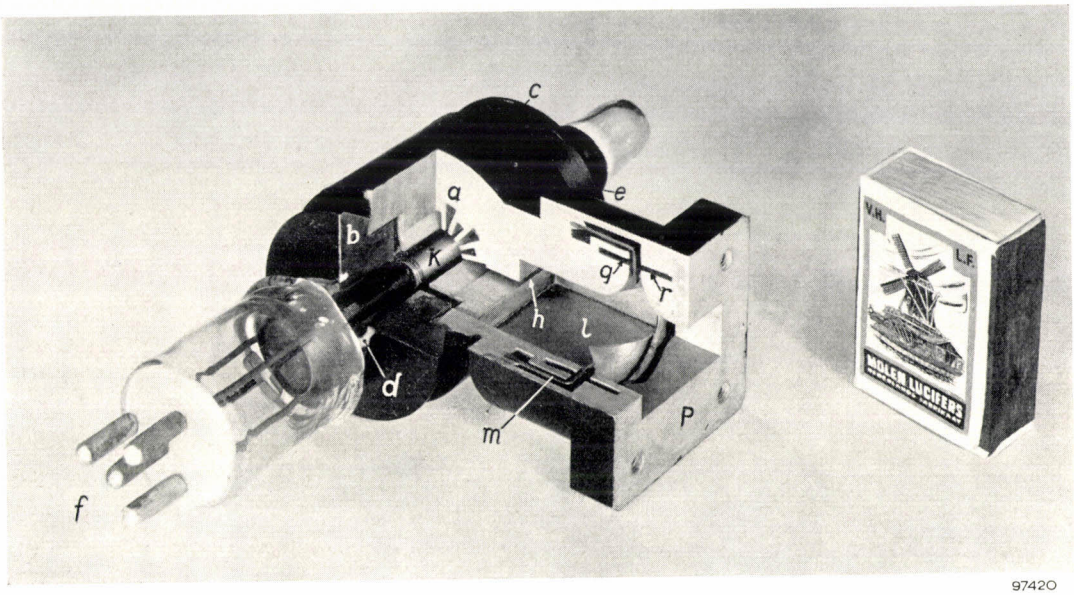
6) The power output remaining unchanged, the dissipation per unit surface rises sharply upon scaling-down. In certain cases this will call for a more elaborate cooling system. With pulsed magnetrons there is always the additional possibility of reducing the *mean* power by using shorter pulses or a lower pulse-repetition frequency.

The foregoing shows that in order to produce an appropriate design it will sometimes be necessary for certain dimensions to differ from those found by scaling-down an existing tube. Examples will be encountered in the following pages.

General construction and some engineering details

The construction will be explained with reference to the cutaway 32 mm magnetron shown in *fig. 3*. The central part is the anode block *a*; the other components project from this along two mutually perpendicular axes.

The anode block is made from a solid copper cylinder, in which nine small and nine large resonant cavities are pressed in the form of sectors (constituting a so-called "rising sun" pattern) by means of a steel hob; the operation is known as hobbing [2]). The eighteen resonant cavities are coupled together primarily via the space between cathode and anode, and together form a resonator system having a series of natural frequencies. In practice the mode of oscillation used is that in which adjacent resonant cavities oscillate in anti-phase ("π mode"); the frequency of this mode is governed by the various dimensions. After hobbing, cylindrical recesses are machined to receive the other components. The first of these components are the two iron end pieces *b* and *c*, which function as pole pieces for the magnet; they fit in the anode block with a small clearance from the ends of the resonant cavities in order to minimize the air-gap for the axial magnetic field. The end pieces are provided with holes through which the cathode is inserted, and also with rings *d* and *e* of fernico, an alloy to which hard glass can be sealed. The glass ring sealed to *d* serves as insulation



97420

Fig. 3. Cutaway view of 32 mm magnetron showing the following components: *a* anode block; *b* and *c* iron end pieces, serving as pole pieces; *d* and *e* fernico rings for hard-glass seals; *k* cathode; *f* connections for cathode and filament (as in fig. 4*a*); *h* transformer slot in copper disc; *l* output waveguide; *m* fernico cap with hole for glass output window; *p* brass connection flange; *q* and *r* slots functioning as RF chokes.

[2]) For further particulars of this process, see G. B. Collins, loc. cit., p. 654-661.

between cathode and anode, and the glass tube sealed to *e* forms the pinch via which the tube is exhausted.

The cathode *k* is an L-type or dispenser cathode, in the form of a porous tungsten cylinder of the same length as the anode block. End shields at the end of this cathode cylinder prevent the axial loss of electrons to the end shields, which would adversely affect the operation of the magnetron. The coiled filament inside the cathode cylinder is fed via the lead-ins through the glass at *f*.

Perpendicular to the axis of cathode and pole pieces is mounted the output system. For this purpose a recess is milled into the anode block so as to remove the rear wall of one of the resonant cavities. Against this a copper disc is fitted provided with a transformer slot *h* which couples the resonant cavity with the waveguide *l*. At the other end of *l* is a vacuum seal consisting of a round window sealed into the fernico cap *m*. (The window is absent in fig. 3.) Finally, the brass flange *p* enables *l* to be joined to the waveguide carrying the energy to the external circuit.

By means of the transformer slot *h* it is possible to adjust the load which the antenna constitutes on the anode system. When the load increases the magnetron efficiency increases also, but at a certain critical load the operation becomes unstable. Moreover, variations in the antenna impedance, such as arise with a rotating antenna, cause changes in the oscillating frequency which are greater the higher the load. It is usually required that, at a reflexion coefficient of 0.2, the maximum frequency drift, termed the pulling figure, should not exceed a specified value. This may involve some compromise in the efficiency. The necessary dimensions of the transformer slot can be determined partly by calculation and partly by experiment.

The window should be made of a type of glass with sufficiently low high-frequency losses, so that there is no danger of it melting at the specified frequency and the specified mean power. Where the mean power is particularly high, intensified air-cooling can be employed. In fig. 3 it can be seen that the waveguide is chamfered where the window fits; this reduces the heat development per cm^2 surface of window. The larger area of window also diminishes the risk of flash-over due to the high RF field strengths, which could easily damage the window. The cylindrical slots *q* and *r* in parts *l* and *p* serve to prevent energy leaking away along the window and the fernico cap; at this frequency they act as RF chokes.

The tube is assembled as follows. The various metal parts are brazed together, the required temperature being produced by high-frequency heating. This is done in a reducing atmosphere, so that the parts remain perfectly clean. After brazing, the glass-to-fernico seals are made and the tube, still without the cathode, is exhausted and tested for leaks. If this test proves satisfactory, the tube is opened up again and the cathode sealed in, special jigs being used to ensure true alignment. The tube is then finally exhausted and degassed, after which the cathode is activated. After sealing-off, the cooling jacket is fitted (not shown in fig. 3) and the coupling flange *p* secured to the assembly with tin solder. The magnetron can then be mounted between the pole pieces of a suitable magnet.

Some constructional details

The foregoing remarks apply equally to all four tubes of the range described. We shall now discuss some special problems that arise in the construction of the individual magnetrons.

High power

With the 32 mm magnetron the aim was to increase both the peak output power and the mean power. A peak power of about 1100 kW was achieved at a mean power of 900 W. The peak power increases with the anode current and the anode voltage; this also entails an increase in the requisite magnetic field. The theoretical limits of current and voltage are difficult to specify; in practice it is found that every magnetron finally becomes unstable as the current and voltage are increased and refuses to oscillate with reasonable efficiency. The maximum peak power is dependent, among other things, on the dimensions of the large and small resonant cavities, the most favourable values for which must be found empirically.

A high mean power calls primarily for good dissipation of the power losses. In the 32 mm magnetron the losses are of the same order of magnitude as the useful power and are dissipated in the anode, the cathode and the output window. The copper anode readily conducts the heat outside, and the use of water cooling permits the dissipation here of several hundreds of watts. If no special precautions are taken, the output window will break down at a useful power of about 500 W, when a few watts are dissipated in the glass. The window can be adequately cooled by passing a stream of air over it as mentioned above.

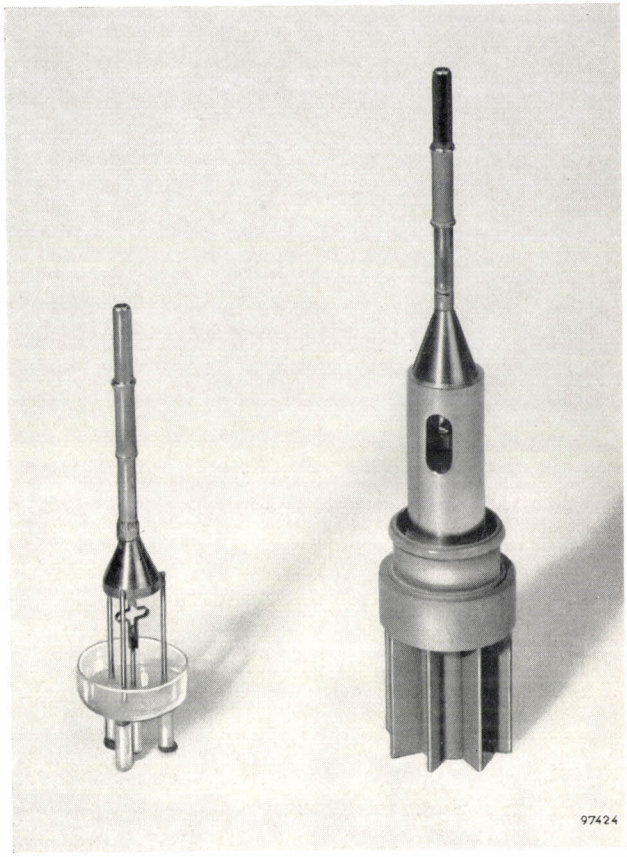
Dissipation in the cathode causes the greatest difficulty. This is due to back bombardment from those electrons in the interaction space that absorb

energy from the high-frequency field. Provided this dissipation is smaller than the heater power required for the cathode, the cathode temperature can be kept constant by reducing the heater current. However, if the back-bombardment power rises above the normal heater power, the cathode, even with zero heater current, becomes overheated, as a result of which the materials used deteriorate more quickly and the life of the cathode is curtailed. In the case of the 32 mm magnetron, several measures were adopted to keep the cathode temperature within reasonable bounds. In the first place the length of the emissive part (and hence the anode length) was made as large as possible without causing undesired resonance modes in the anode system. A longer system means a smaller cathode dissipation per cm^2. Furthermore the heat dissipation of the cathode was made as high as possible — just the opposite to what is done for the cathodes in other tubes. The rate of heat loss can be increased in the first place by improving the heat conduction. For this purpose thicker and better-conducting materials can be used for the cathode support and leads. Secondly one can attempt to increase the heat loss by radiation, which is considerable at the hottest part of the cathode surface. With the cathode in question this was done by giving the porous tungsten cylinder the roughest possible surface. *Fig. 4* shows side by side a cathode (*a*) for which no special measures have been taken to improve the heat dissipation, and a cathode (*b*) designed to give maximum heat dissipation. At an emitting-surface temperature of 1200 °C the heat dissipation was found to be 90 W at cathode *a* and 120 W at cathode *b*. The latter is the normal cathode dissipation for this tube at a mean power of 900 W. With the given dimensions of the emitting surface this probably represents the maximum heat dissipation obtainable in this way.

At these high values of mean power the L cathode is clearly superior to the oxide cathode, in view of the higher temperature which the L cathode needs for normal emission. Moreover an L cathode is much better able than an oxide cathode to withstand temperatures higher than the normal operating temperature.

High frequencies

We shall now discuss some of the difficulties involved in the scaling-down of dimensions by taking as an illustration the smallest magnetron of the range, the 4 mm type. The anode aperture is 1.4 mm in diameter. Around it are grouped 18 resonant cavities, separated one from the other by copper vanes 0.14 mm thick — precision work, beside which the mechanism of many a wristwatch looks positively coarse. Obviously the grinding of the hob for such a miniature system is no easy matter. Without going into details it may be mentioned that one aspect of the utmost importance is to keep the machined



a *b*

Fig. 4. Cathodes for the 32 mm magnetron. Model *a* has a heat dissipation of 90 W, model *b* of 120 W.

metal, the support and the grinding tool at a constant temperature. A good hob is capable of making hundreds of anode systems.

The cathode, illustrated in *fig. 5*, has an emitting portion of 0.8 mm diameter. The greatest difficulty here was centring the cathode in the anode aperture, the distance between cathode and anode being only 0.3 mm. Although the cathode is very accurately sealed to the correct position by means of jigs, during the subsequent cooling the very slight deformation due to stresses in the glass and cathode components is still enough to cause impermissible eccentricity. This did in fact lead to many rejects. For this reason a special construction was adopted, as shown schematically in fig. 5. The anode block *a* with end pieces *b* and *c* is joined via the glass cylinder *d* to the cathode connection *e*. This part is assembled first, care being taken to align the hole of the collar

in *e* very accurately with the anode aperture. The cathode *k* is then introduced and centred in the anode system with the aid of a jig. This is made possible by a small clearance between the cathode and the collar in *e*. These parts are then soldered together at *g*. As a rule soldering causes much less deformation than the sealing process, and therefore good results can be obtained with this method. The final operation in this construction is to seal-in the filament pin *f*.

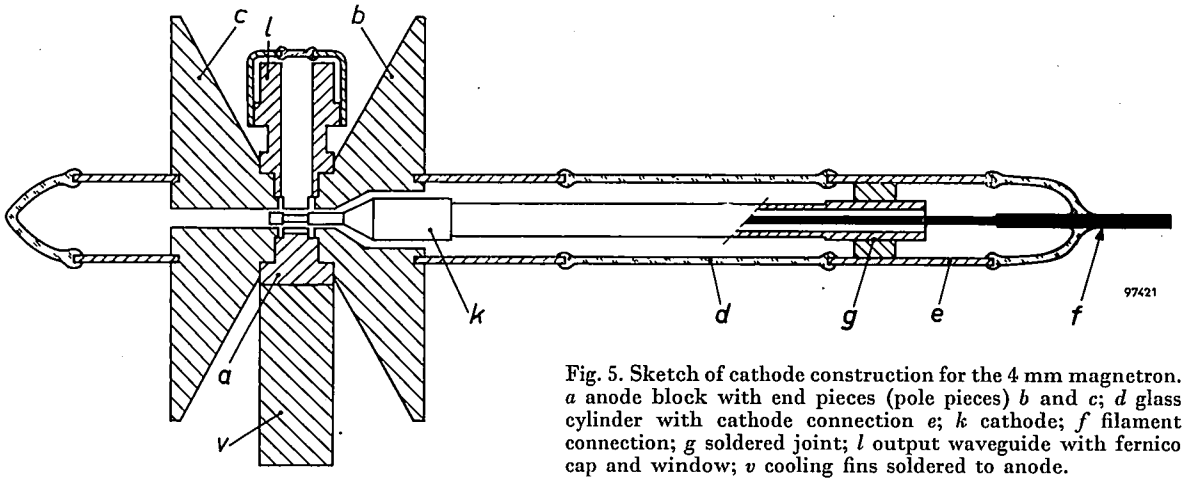


Fig. 5. Sketch of cathode construction for the 4 mm magnetron. *a* anode block with end pieces (pole pieces) *b* and *c*; *d* glass cylinder with cathode connection *e*; *k* cathode; *f* filament connection; *g* soldered joint; *l* output waveguide with fernico cap and window; *v* cooling fins soldered to anode.

The current density at the surface of the cathode in the 4 mm magnetron is about 200 A/cm². This is appreciably more than the saturation emission possible at reasonable cathode temperatures; the extra is supplied by the secondary emission due to back bombardment, which thus plays an essential role in this tube.

Comparison of results with the scaling factors derived from similarity considerations

The most important dimensions of a magnetron are denoted by letters in *fig. 6*. Column *a* in *Table I* gives the values of these dimensions and the exact wavelength in millimetres for each tube. To facilitate comparison, the dimensions are also expressed for each tube in terms of its wavelength (reduced dimensions, column *b*).

The table shows that, except for the anode length, the systems are in fact geometrically similar within narrow limits and that the measured wavelength is proportional to the linear dimensions. The reduced anode length in the 32 mm and 4 mm magnetrons is about twice that in the others. In the 32 mm magnetron this was done to obtain the highest possible mean power, the cathode dissipation per unit surface being — as we have seen — the chief limiting factor. In regard to the 4 mm magnetron the consideration was to obtain a high total operating current in spite of the limited emission current of the cathode. The advantage of changing the anode length in this connection is that, fundamentally, it has little effect on the operation, since, theoretically at least, no currents flow in the axial direction. (It is again the practical deviations from this that set a limit to the increase of this dimension.)

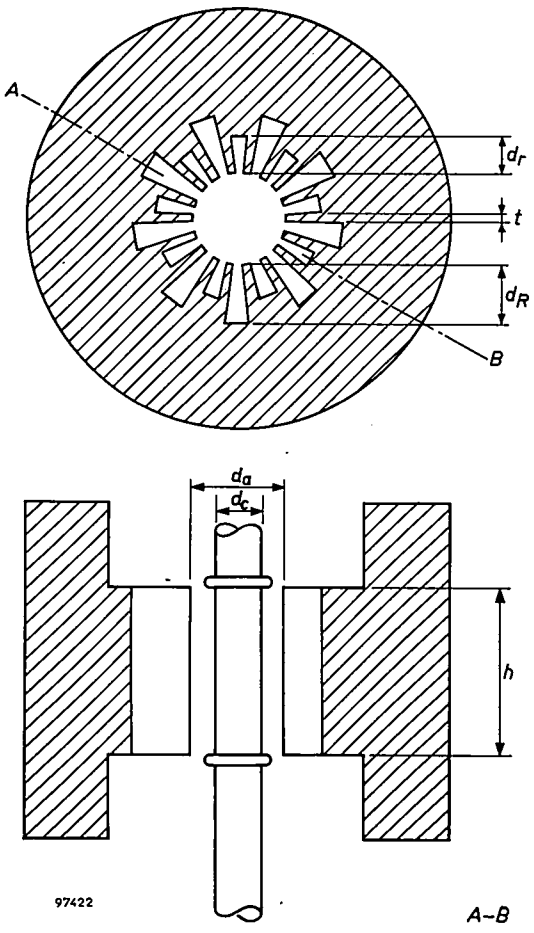


Fig. 6. Simplified cross-section of the magnetrons described here, indicating the principal dimensions: d_a = anode diameter, d_c = cathode diameter, h = anode length, d_r = depth of small resonant cavities, d_R = depth of large resonant cavities, and t = thickness of vane.

The principal characteristics of the magnetrons under discussion are represented in the performance charts in *fig. 7a, b, c* and *d*, in which contours of constant magnetic induction B, of constant power output P and of constant efficiency η are plotted with the anode current I_a and the anode voltage V_a as coordinates.

Table I. The principal dimensions (see fig. 6) of the four magnetrons, a in mm, b in terms of the wavelength as unit of length.

Dimension \ Magnetron	32 mm		12 mm		8 mm		4 mm	
	a	b	a	b	a	b	a	b
Wavelength λ	31.5		12.2		8.50		3.97	
Anode diameter d_a	10.6	0.34	4.05	0.33	2.90	0.34	1.40	0.35
Cathode diameter d_c	6.5	0.21	2.45	0.20	1.85	0.22	0.80	0.20
Anode length h	21.0	0.67	3.85	0.32	2.72	0.32	2.50	0.63
Depth of small resonant cavities d_r	4.10	0.13	1.59	0.13	1.16	0.14	0.56	0.14
Depth of large resonant cavities d_R	7.38	0.23	2.92	0.24	2.08	0.24	0.94	0.24
Thickness of vane t	1.12	0.036	0.43	0.035	0.308	0.036	0.143	0.036

Comparison of the charts for the four magnetrons reveals clearly the way in which a design according to the scaling law influences the characteristics. In general the maximum voltage and the maximum current are decreased by scaling-down, the first particularly because of the impossibility of generating sufficiently strong magnetic fields, and the second because of the limited emission current density of the cathode. To illustrate this the operating region of the 4 mm magnetron is shown shaded in all charts; strictly, the four performance charts can only be compared with each other in this region.

For a numerical comparison we may select two arbitrary points in the chart, namely those having the coordinates (6 A, 15 kV) and (4 A, 13 kV). If we consider these points for the 8 mm and 12 mm magnetrons, we see that, from similarity considerations, the corresponding points for the 32 mm and 4 mm magnetrons are those with twice the current, the anodes being relatively twice as long, i.e. the points (12 A, 15 kV) and (8 A, 13 kV), respectively. The values of the characteristics read from the charts at these points are given in *Table II*, together with the product $B\lambda$. In the η column we see that, with the exception of the 12 mm magnetron (see below), the efficiency decreases monotonically with the wavelength. This is due to the fact that the conductance decreases as the wavelength shortens (owing to skin effect), whereas the scaling factor requires it to increase. Measurements of the Q of the resonator system have shown that an efficiency drop of this order of magnitude is indeed to be expected.

The exception in the case of the 12 mm magnetron is due to the coupling to the waveguide being much looser, so that in this respect it is not a scaled-down model. As mentioned, this reduces efficiency, but its advantage is that it increases the stability, as appears from the pulling figure, which is here 16 Mc/s against 13 Mc/s with the 32 mm type and 32 Mc/s with the 8 mm type. (It will be recalled that the ratio of oscillating frequency to pulling figure is a measure of a magnetron's stability.)

The last column gives the product $B\lambda$ which, from similarity considerations, should remain unchanged when the dimensions are scaled-down. The table shows that this is true only to a first approximation. Singling out the 12 mm magnetron, we notice that this product decreases towards shorter wavelengths. The reason is to be found in the limited primary emission of the cathode; conditions of exact similarity do not therefore obtain with this range of magnetrons. In the 32 mm magnetron the saturation emission is not quite reached, and so the electric field at the cathode is virtually zero. The primary emission in the 4 mm magnetron, on the other hand, is saturated (see above), so that strong electric fields appear at the cathode. If, however, the cathode temperature in the 8 mm and 4 mm magnetrons is increased, the anode voltage and current

Table II. Comparison of characteristic values for the four magnetrons, read from the performance charts (fig. 7*a, b, c, d*) at corresponding points according to similarity considerations. The comparison is made for two sets of points.

Wavelength λ (mm)	I_a (A)	V_a (kV)	B (Wb/m²)	P (kW)	η (%)	$B\lambda$ (10^{-4} Wb/m)
31.5	12	15	0.28	54	30	90
12.2	6	15	0.83	14	16	101
8.50	6	15	1.03	23	25	88
3.97	12	15	1.85	34	19	73
31.5	8	13	0.26	30	29	83
12.2	4	13	0.76	8	15	92
8.50	4	13	0.90	8	15	76
3.97	8	13	1.6	12.5	12	64

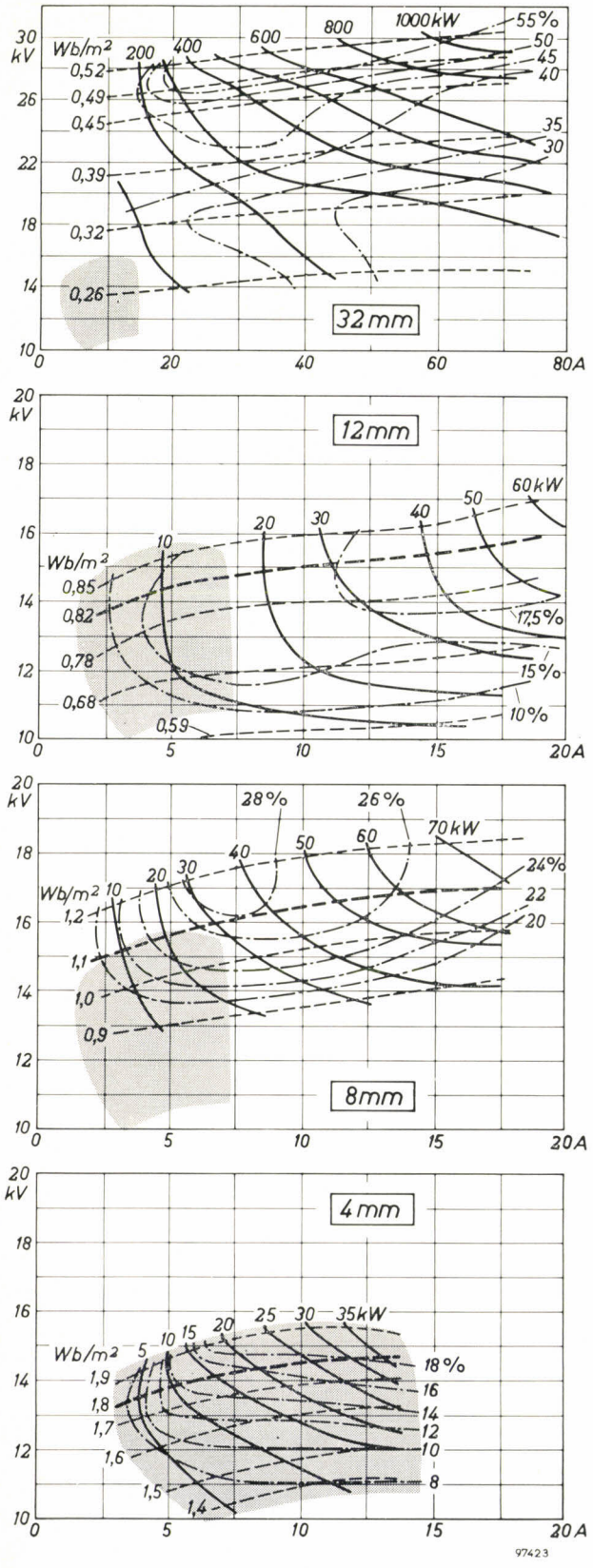


Fig. 7. Performance charts for the four magnetrons: *a*) 32 mm, *b*) 12 mm, *c*) 8 mm and *d*) 4 mm. On the abscissa the anode current I_a, on the ordinate the anode voltage V_a.
— — — — — Contours of constant magnetic induction B (the thicker contour in *b*, *c* and *d* relates to the field produced by the permanent magnet).
———— Contours of constant power P.
—·—·—·—·— Contours of constant efficiency η.

remaining the same, stronger magnetic fields are indeed required, i.e. the product $B\lambda$ is closer to the value expected from similarity considerations.

Life tests

The life tests on the magnetrons demonstrate the excellent properties of the L cathode. Not one of these tests was terminated by lowered emission. The most frequent cause of failure was leakage due to inadequate cooling; the 8 mm magnetron was still working well after a life test of almost 1500 hours. No life test has yet been made on the 4 mm magnetron. The results of the tests are given in *Table III*. The

Table III. Particulars of life tests on a 32 mm, a 12 mm and an 8 mm magnetron.

Wavelength in mm	31.3	12.2	8.35
Anode current in A	55	14	14
Anode voltage in kV	31	15.4	16.8
Peak power in kW at beginning of test	783	46	52
Peak power in kW at end of test	695	40	45
Efficiency in % at beginning of test	46	21	22
Efficiency in % at end of test	41	18	19
Pulse duration in μsec	2.0	0.44	0.50
Pulse-repetition frequency in c/s	500	2300	1000
Mean initial power in W	783	46	26
Duration of test in hours	238	700	1488

pulse durations and the repetition frequencies were rather arbitrarily chosen; the mean powers given for the 12 mm and 8 mm magnetrons are therefore not the maximum permissible values. A point of importance for short-range radar is that the 8 mm and 4 mm magnetrons work excellently with pulses of only 0.01 μsec.

Summary. An experimental range of magnetrons is described for wavelengths of 32, 12, 8 and 4 mm, and delivering peak outputs of 1100, 70, 80 and 40 kW, respectively. Since the four magnetrons have virtually the same geometrical proportions, the laws of similarity are theoretically applicable to them. It is found that the wavelength is proportional, and the requisite magnetic field inversely proportional, to the linear dimensions and that the same power is generated for the same anode voltage and current. Towards shorter wavelengths, however, difficulties arise owing *inter alia* to increasing high-frequency resistance, the limitation of the attainable magnetic field and cathode emission. Some constructional details are discussed. The "rising sun" anode system is hobbed from a solid copper block; to this are brazed iron end pieces (pole pieces). The coupling slot, the output waveguide and its glass window seal are also discussed. The axially mounted cathode is of the L type. Some special constructional problems are touched on with reference to the 32 mm and 4 mm magnetrons. The first was designed for a high mean power (900 W), which necessitated a cathode construction permitting a high rate of heat dissipation; the small dimensions of the 4 mm magnetron called for a special method of assembling the cathode in the tube. Comparison of the results of the various tubes shows that the wavelength is closely proportional to the linear dimensions and that the mutual disparities in efficiency can be satisfactorily explained. Life tests on a 32 mm, a 12 mm and an 8 mm magnetron were terminated after 238, 700 and 1488 hours, respectively; the power outputs had dropped in that time by about 15%.

THE RESISTANCE NETWORK, A SIMPLE AND ACCURATE AID TO THE SOLUTION OF POTENTIAL PROBLEMS

by J. C. FRANCKEN.

518.5:53.072.13:621.317.729

Among the methods that can be employed to solve the Laplace equation for given boundary conditions, that involving the use of a resistance network is in many cases highly attractive. It is a method applicable to two-dimensional problems and to three-dimensional problems where there is rotational symmetry. Setting-up the boundary conditions is particularly easy, the measurements do not take up much time, and remarkably accurate results are attainable.

In many branches of physics one is frequently confronted with potential problems, the solution of which involves finding a function φ which satisfies the Laplace equation, viz., in rectangular coordinates,

$$\frac{\partial^2\varphi}{\partial x^2}+\frac{\partial^2\varphi}{\partial y^2}+\frac{\partial^2\varphi}{\partial z^2}=0. \quad . \quad . \quad . \quad (1)$$

Examples of quantities satisfying the equation are the electrical potential in a space-charge free region and the gravitational potential in the space between the gravitating masses. The temperature under steady-state conditions of heat flow, and velocity potential in a non-turbulent stream of incompressible fluid are further examples. Usually the value of the potential function φ on certain closed surfaces, e.g. at solid boundaries, is known. In addition, sometimes the space in which φ must satisfy (1) is entirely enclosed by a surface at which φ is known. In other cases the space extends to infinity, at which φ approaches a known constant value. If φ represents temperature, the constant value will be the ambient temperature; if it represents electric potential, the constant value will be earth potential. The example we shall be using to illustrate the employment of the resistance network is drawn from electron optics, and concerns the potential distribution in electron guns for television picture-tubes. The surfaces where φ has known values will be those of the electrodes of the tube.

Only in a few simple cases is it possible to express the required potential function explicitly in terms of its given boundary values at certain surfaces. Usually one has to proceed by other methods. Apart from numerical methods, which are now widely employed in conjunction with digital computers, analogue techniques are particularly suited to potential problems. One well-known analogue technique involves the use of an electrolytic tank. A model of the electrode assembly, often an enlarged model, is submerged in a conducting liquid. The potential distribution existing when given voltages are applied to the electrodes is not altered when the model is submerged. The potential distribution in the electrolyte is measured with a probe [1)][2)][3)].

The electrolytic tank has proved to be a valuable aid in the solution of electron-optical problems. It has, however, its limitations and drawbacks: the construction of the electrode models is often laborious and expensive, and measurements in their vicinity are inaccurate because the liquid near the electrodes rises in consequence of capillarity. For three-dimensional problems with rotational symmetry tanks of "wedge" section are often used. A vertical section through this tank has the shape of a wedge. Electrodes can often be constructed from strips of metal, the disadvantages of complicated models thus being avoided. On the other hand a new disadvantage arises in that measurement near the axis of symmetry, precisely the most important region, is rendered very inaccurate by the marked capillary rise at the sloping bottom of the tank. It is not therefore surprising that, apart from the electrolytic tank, other analogue techniques have been developed for the purpose of determining potential distributions. One of these is the resistance network, which is the subject of the present article[4)].

---

1) G. Hepp, Measurements of potential by means of the electrolytic tank, Philips tech. Rev. 4, 223-230, 1939.

2) N. Warmoltz, Potential distribution at the igniter of a relay valve with mercury cathode, Philips tech. Rev. 8, 346-352, 1946.

3) An example of the use of the electrolytic tank for determining temperature distribution is described in F. Reiniger, The study of thermal conductivity problems by means of the electrolytic tank, Philips tech. Rev. 18, 52-60, 1956/57.

4) Another method involves the use of conducting (graphite-surfaced) paper. This method can be employed for solving two-dimensional problems. Electrodes are simulated by cutting them out in copper foil. Conducting paper cannot be used for solving rotationally-symmetric three-dimensional problems, whereas the resistance network can be used in such cases.

In theory, the solutions it provides are only approximate; nevertheless in many cases these solutions are much more exact than those obtained with the aid of the tank. The resistance network was first proposed by Hogan [5]) and first used by De Packh [6]). However, the credit for pointing out the high degree of accuracy thereby attainable belongs to Liebmann [7])[8]).

Explanation of the principle of the network becomes more straightforward if the terms "differential operator" and "finite-difference operator" can be employed. We therefore start by giving a brief account of the two operators. Resistance networks for two-dimensional and for rotationally-symmetric three-dimensional problems are then discussed, followed by a description of an actual network of the latter kind, and of the manner in which it is used. This will be illustrated with the aid of a practical application. Finally, the implications of the fact that the network has finite meshes will be discussed with reference to two examples.

Differential and difference operators

The sum of the second differential coefficients with respect to x, y and z of any function $w(x,y,z)$ is often denoted by the symbol ∇^2:

$$\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}+\frac{\partial^2 w}{\partial z^2}\equiv\nabla^2 w. \quad . . \quad (2)$$

∇^2 is called the Laplacian operator. The effect of the operator on the function w will naturally depend on the point in space at which it is applied (i.e. $\nabla^2 w$ is again a function of x, y and z). We are now able to express the Laplace equation (i.e. (1) above) in words as follows: *a function satisfying the equation has the property that when operated on by the Laplacian ∇^2, the result is zero at every point.*

From the differential operator ∇^2 we shall now derive a finite-difference operator L that, operating on the same function at the same point, produces almost the same effect as ∇^2 itself, i.e. $Lw \approx \nabla^2 w$. For this purpose we consider an arbitrary point O and three pairs of points, P and Q, R and S, and T and U, which lie respectively in the x, y and z directions at a distance a on either side of O (*fig. 1*).

We can express the difference between the values w_Q and w_O of w at points Q and O by using Taylor's theorem. This gives us the difference in terms of a and the derivatives of w with respect to x at point O:

$$w_Q-w_O=a\left(\frac{\partial w}{\partial x}\right)_O+\frac{a^2}{2!}\left(\frac{\partial^2 w}{\partial x^2}\right)_O+ +\frac{a^3}{3!}\left(\frac{\partial^3 w}{\partial x^3}\right)_O+\frac{a^4}{4!}\left(\frac{\partial^4 w}{\partial x^4}\right)_O+\ldots . \quad (3a)$$

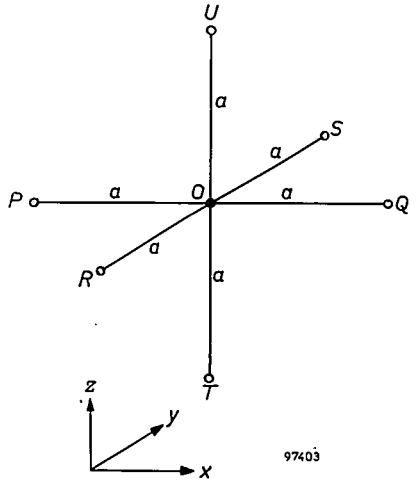


Fig. 1. Illustrating the derivation of the difference operator L.

By replacing a in the above by $-a$, we obtain a similar series for w_P-w_O:

$$w_P-w_O=-a\left(\frac{\partial w}{\partial x}\right)_O+\frac{a^2}{2!}\left(\frac{\partial^2 w}{\partial x^2}\right)_O- -\frac{a^3}{3!}\left(\frac{\partial^3 w}{\partial x^3}\right)_O+\frac{a^4}{4!}\left(\frac{\partial^4 w}{\partial x^4}\right)_O-\ldots . \quad (3b)$$

Adding these two series and solving for $(\partial^2 w/\partial x^2)_O$, we obtain:

$$\left(\frac{\partial^2 w}{\partial x^2}\right)_O=\frac{1}{a^2}\{(w_Q-w_O)+(w_P-w_O)\}- -\frac{a^2}{12}\left(\frac{\partial^4 w}{\partial x^4}\right)_O-\ldots . \quad (4)$$

The differential coefficient $(\partial^2 w/\partial x^2)_O$ is thus expressed in terms of the differences (w_Q-w_O) and (w_P-w_O) plus a number of correction terms, whose total value can be made as small as desired simply by making a small enough.

$(\partial^2 w/\partial y^2)_O$ and $(\partial^2 w/\partial z^2)_O$, the other differential coefficients occurring in (2) above, can be expressed as differences in an analogous manner. Inserting in (2) the expressions thus obtained, we have:

---

[5]) T. K. Hogan, A general experimental solution of Poisson's equation for two independent variables, J. Instn. Engrs. Austr. **15**, 89-92, 1943.

[6]) D. C. de Packh, A resistor network for the approximate solution of the Laplace equation, Rev. sci. Instr. **18**, 798-799, 1947.

[7]) G. Liebmann, Solution of partial differential equations with a resistance network analogue, Brit. J. appl. Phys. **1**, 92-103, 1950.

[8]) G. Liebmann, Field plotting and ray tracing in electron optics, Advances in electronics **2**, 101-149, 1950.

$$(\nabla^2 w)_0 = \frac{1}{a^2}(w_P + w_Q + w_R + w_S + w_T + w_U - 6w_0) - \\ - \frac{a^2}{12}\left(\frac{\partial^4 w}{\partial x^4} + \frac{\partial^4 w}{\partial y^4} + \frac{\partial^4 w}{\partial z^4}\right)_0 - \ldots\ldots \quad (5)$$

Introducing the operator L, we now put:

$$(Lw)_0 = \frac{1}{a^2}(w_P + w_Q + w_R + w_S + w_T + w_U - 6w_0). \quad \ldots \quad (6)$$

We see from (5) that $(Lw)_0$ is an approximation to $(\nabla^2 w)_0$, approaching it all the more closely according as a is made smaller. L is the *finite-difference operator* referred to above; it is so called because L denotes an operation whereby the finite differences $w_P - w_0$, etc. are used.

This procedure for deriving a difference operator from a differential operator can also be followed in the more general case where the differential equation involves the first derivatives ($\partial w/\partial x$, etc.) as well as the second derivatives. By subtracting equations (3a) and (3b) from each other, one can arrive at an expression for $(\partial w/\partial x)_0$ involving $(w_P - w_0)$ and $(w_Q - w_0)$ and a series of correction terms whose sum approaches zero as a goes to zero. We shall make use of this when dealing with three-dimensional problems with rotational symmetry.

In the special case where w is independent of z, we have $w_T = w_U = w_0$, and (6) simplifies to

$$(Lw)_0 = \frac{1}{a^2}(w_P + w_Q + w_R + w_S - 4w_0). \quad (7)$$

We can now go on to discuss the principle of the resistance network.

A resistance network for two-dimensional problems

To start with we shall confine ourselves to two-dimensional cases. There is then one direction — the z-direction say — in which the function being sought does not vary. In these cases $\partial^2\varphi/\partial z^2 = 0$, and Laplace's equation assumes the form:

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = 0. \quad \ldots\ldots \quad (8)$$

To take a definite case, let us consider the example shown in *fig. 2*. The three closed outlines s_1, s_2 and s_3 represent sections taken at right angles through three infinitely long prisms. On the periphery of each prism φ has a known constant value. The problem is to find a function φ which satisfies (8) in the area within s_3 but outside s_1 and s_2, and which assumes the prescribed values along s_1, s_2 and s_3.

Over s_1, s_2 and s_3 we place a square grid, the lines of which are parallel to the x and y axes and spaced at intervals of a (see fig. 2). The "grid lines" intersect at "grid points". We shall refer to two grid points as "adjacent" if their distance apart is the mesh width a. In fig. 2 the outlines s_1, s_2 and s_3 have been drawn in such a way that they intersect the grid only in grid points. This is not a necessary limitation, but it will simplify discussion, and the examples we shall be dealing with will subject to it.

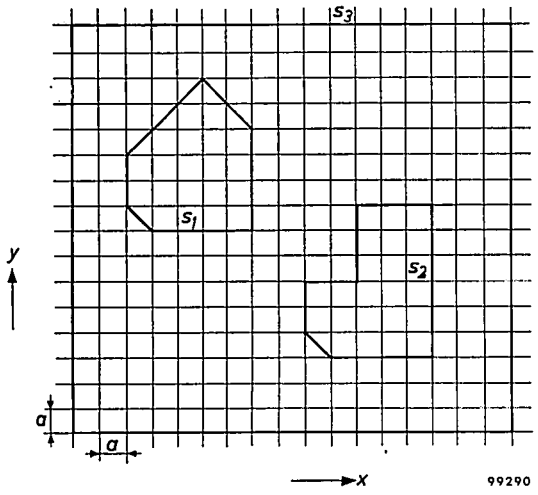


Fig. 2. The unknown function φ has known values along the outlines s_1, s_2 and s_3. Throughout the area inside s_3 but outside s_1 and s_2 it must satisfy the Laplace equation. It is assumed that s_1, s_2 and s_3 are such that they intersect a superposed square grid (mesh width a) only in grid points.

We shall refer to grid points located on the outlines s_1, s_2 and s_3 as "boundary grid points", and to the remaining ones in the area wherein φ has to be determined as "internal grid points".

The following proposition underlies the principle of the resistance network employed.

If each internal grid point is allotted a value φ^* such that between the value φ^* at any such point and the values at adjacent grid points the relationship $L\varphi^* = 0$ exists and if at the boundary grid points φ^* has the boundary values specified for the required function φ at those places, then the difference between φ^* and φ at the internal grid points will approach zero as the mesh width (a) approaches zero.

As we have already pointed out, there is a close connection between the difference operator L and the differential operator ∇^2; the proposition just stated is accordingly a plausible one. Its truth can be proved rigorously by demonstrating that at all grid points $\varphi - \varphi^*$ is smaller than a certain finite

quantity that approaches zero as the mesh width a does so [9]).

For the purpose of finding φ^* values, a network of resistors is built up. We shall see later what requirements the component resistors have to satisfy. The junctions of the resistance network will correspond to the grid points in fig. 2; accordingly, four resistors will meet at each junction (*fig. 3*). We shall refer to junctions corresponding to boundary grid points as "boundary junctions". Between the boundary junctions we may apply voltages that are proportional to the differences between φ values in the corresponding boundary grid points. If the lowest value of φ at any of the boundary grid points is $\varphi_{\min}$ and if we take the potential of the corresponding boundary junction as a datum for measuring the potentials V_b of other boundary junctions, then any of these latter potentials is given by

$$V_b = (\varphi_b - \varphi_{\min})/\beta. \quad . \quad . \quad . \quad . \quad . \quad (9)$$

The suffix b to V and φ indicates that they relate to *boundary* junctions in the resistance network and to the corresponding boundary grid points. $1/\beta$ is a constant of proportionality. We now allot to each *internal* grid point a value

$$\varphi^* = \beta V + \varphi_{\min}, \quad . \quad . \quad . \quad . \quad (10)$$

where V is the potential measured at the corresponding junction in the resistance network. Through a resistor connecting two adjacent junctions in the network, e.g. that between P and O in fig. 3, flows a current having the value $(V_P - V_O)/R_1$. Applying Kirchhoff's law to junction O, we find that

$$\frac{V_P - V_O}{R_1} + \frac{V_Q - V_O}{R_2} + \frac{V_R - V_O}{R_3} + \frac{V_S - V_O}{R_4} = 0.$$

From this and from (10) it follows that for any grid point O,

$$\frac{\varphi_P^* - \varphi_O^*}{R_1} + \frac{\varphi_Q^* - \varphi_O^*}{R_2} + \frac{\varphi_R^* - \varphi_O^*}{R_3} + \frac{\varphi_S^* - \varphi_O^*}{R_4} = 0. \quad . \quad (11)$$

Comparison of (11) with (7) makes it clear that, *if all four resistors have the same value,*

$$(L\varphi^*)_O = 0$$

at any grid point O. Since, in addition, φ^* on the boundary curves has the boundary values laid down for φ, φ^* constitutes an approximation to the required function φ, provided all resistors composing the network are of the same value. The question as to how close the approximation is will be dealt with at the end of the article.

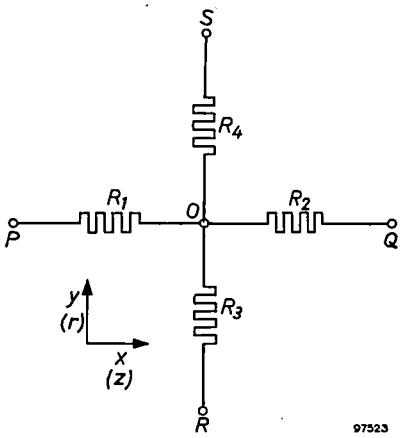


Fig. 3. Resistors R_1, R_2, R_3 and R_4 meet at O, a junction in the resistance network. In the two-dimensional case, PQ represents a line parallel to the x-axis, and RS a line parallel to the y-axis; in the rotationally-symmetric three-dimensional case PQ represents a line parallel to the z-axis, RS one parallel to the r-axis.

Three-dimensional problems with rotational symmetry

More common than two-dimensional problems are three-dimensional problems with rotational symmetry. If the rectangular coordinates are converted to cylindrical coordinates (r, z and ϑ in *fig. 4*), the z-axis being made to coincide with the axis of symmetry, the Laplace equation assumes the form:

$$\frac{\partial^2\varphi}{\partial r^2} + \frac{1}{r}\frac{\partial\varphi}{\partial r} + \frac{\partial^2\varphi}{\partial z^2} = 0. \quad . \quad . \quad (12)$$

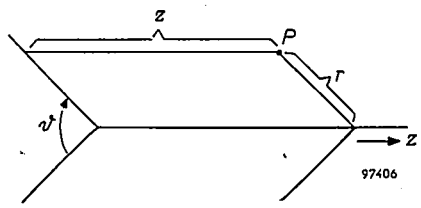


Fig. 4. r, z and ϑ, cylindrical polar coordinates of a point P.

[9]) See S. Gerschgorin, Z. angew. Math. Mech. **10**, 373-382, 1930, or L. Collatz, Numerische Behandlung von Differentialgleichungen, Springer, Berlin, 2nd edition, 1955, p. 320 et seq. Obtaining numerical solutions of the Laplace equation likewise often involves the solution (preferably with the aid of a computer) of the finite-difference equation $L\varphi^* = 0$, and the results obtained are accordingly only approximate. However, these methods can be based on difference operators that are better approximations to the Laplacian than L is, the series of correction terms in (5) beginning with a higher power of a than a^2. If terms involving a^2 etc. are to be eliminated, additional equations must of course be available; this means considering not only adjacent grid points, but also additional grid points in the vicinity of a given grid point. See for example Collatz, loc. cit., p. 352.

Owing to the rotational symmetry ϑ, does not appear in the equation. As in the two-dimensional case, therefore, we have a differential equation involving two independent variables, these being z and r instead of x and y. The Laplacian operator is now

$$\left(\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{\partial^2}{\partial z^2}\right). \quad . \quad . \quad . \quad . \quad (13)$$

In order to derive a finite-difference operator from (13), we let (13) operate on an arbitrary function $u(z,r)$, and consider a point O and the two pairs of points P,Q and R,S which lie in the z and r directions respectively, at a distance a on opposite sides of O. Expressing the differential coefficients as differences, in the same manner as on pp. 11 and 12, we arrive at the following:

$$\left(\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\frac{\partial^2 u}{\partial z^2}\right)_O = (\mathrm{M}u)_O -$$

$$-\frac{a^2}{12}\left(\frac{\partial^4 u}{\partial r^4}+\frac{2}{r}\frac{\partial^3 u}{\partial r^3}+\frac{\partial^4 u}{\partial z^4}\right)_O - \ldots, \quad . \quad . \quad . \quad (14)$$

where

$$(\mathrm{M}u)_O = \frac{1}{a^2}\left\{(u_P-u_O)+(u_Q-u_O)+\right.$$

$$\left.+\left(1-\frac{a}{2r}\right)(u_R-u_O)+\left(1+\frac{a}{2r}\right)(u_S-u_O)\right\}. \quad (15)$$

The required finite-difference operator M is thus defined by (15).

We now superpose on the z,r-plane a square grid with a mesh width of a. The mathematical proposition stated for L on p. 12 is also valid for M [10]). The values of φ^* that are allotted to the internal grid points must now satisfy the relationship

$$\mathrm{M}\varphi^* = 0, \quad (16)$$

while φ^* at the boundary grid points must have the boundary values laid down for φ. φ^* will then be an approximation to φ; in other words, φ^* will approach φ as the mesh width a approaches zero.

As before, it is possible to build up a resistance network whose junctions have potentials corresponding to φ^* values. In order to deduce the requirements the resistors will have to satisfy, we shall proceed as before, but this time (11) must be compared with (15). Having done this, we find that φ^* satisfies the relationship $(\mathrm{M}\varphi^*)_O = 0$ provided that

$$\frac{1}{R_1}:\frac{1}{R_2}:\frac{1}{R_3}:\frac{1}{R_4} = 1:1:\left(1-\frac{a}{2r}\right):\left(1+\frac{a}{2r}\right) \quad (17)$$

(see fig. 3).

If the grid is so positioned that the z-axis coincides with one of the grid lines, then at each of the grid points

$$r = ja,$$

j being an integer. (17) now becomes:

$$\frac{1}{R_1}:\frac{1}{R_2}:\frac{1}{R_3}:\frac{1}{R_4} = 2j:2j:(2j-1):(2j+1). \quad (18)$$

In the present case, then, the resistance values must decrease with increasing distance from the z-axis. In the network shown in *fig. 5*, (18) is satisfied for all values of j except $j=0$. This exception is a point that we must look into.

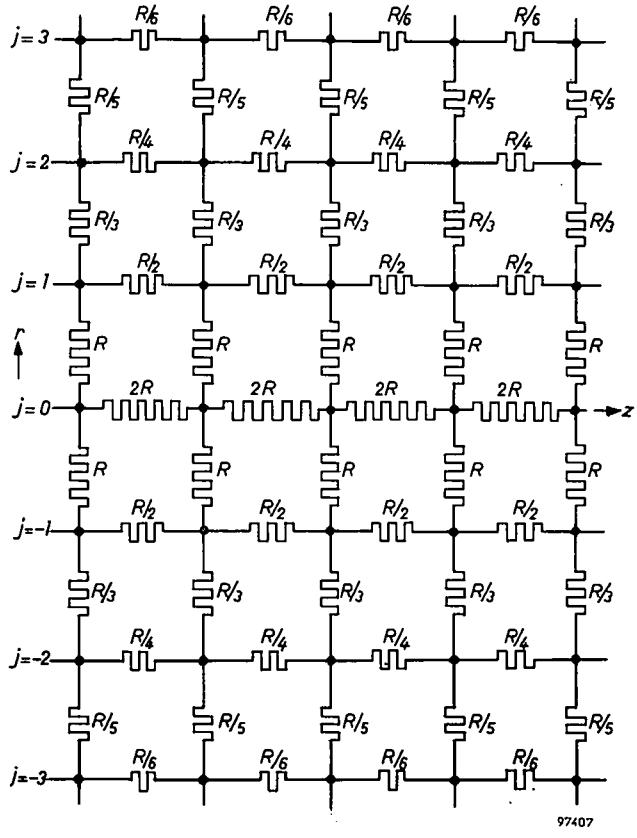


Fig. 5. Part of a network for rotationally-symmetric three-dimensional problems. Resistors meeting at junctions at which $j \neq 0$ have values satisfying the relation (18). For $j = 0$ (junctions on the axis) the resistors satisfy (20).

The resistors lying along the axis

The zero value of j, appropriate to points on the axis of symmetry (z-axis), gives rise to complications. This is clear for example from the conclusion

[10]) For the proof, see Gerschgorin, loc. cit. The proof given there is a general one valid for all elliptical differential equations. (8) and (12) are equations of this type.

that can be drawn from (18), namely that for resistors meeting at a junction on the axis, R_3 must be negative if R_1, R_2 and R_4 are positive. In addition, (14) and (15) involve indeterminate 0/0 terms when $r = 0$ [11]). These complications can be avoided by reverting to the finite-difference operator L (see (6)) as an approximation to the Laplace operator for points along the z-axis. The reason we abandoned L in favour of operator M when we took up the three-dimensional case with rotational symmetry was that the former would have led to a three-dimensional network. This objection does not, however, apply to points on the axis of symmetry. Let us consider a point O on that axis (see fig. 1; we shall assume that the x-axis in this figure represents the axis of symmetry). At point O, then,

$$w_T = w_R = w_U = w_S .$$

For such a point, therefore, we can rewrite (6) in the form

$$(\mathrm{L}w)_O = \frac{1}{a^2}\{(w_P - w_O) + (w_Q - w_O) + 2(w_R - w_O) + 2(w_S - w_O)\} . \quad . \quad . \quad . \quad (19)$$

Comparison of (11) with the above expression makes it clear that, in the grid points on the z-axis, φ^* will satisfy $\mathrm{L}\varphi^* = 0$ provided that

$$\frac{1}{R_1} : \frac{1}{R_2} : \frac{1}{R_3} : \frac{1}{R_4} = 1 : 1 : 2 : 2 . \quad . \quad . \quad (20)$$

These conditions have in fact been satisfied in the network of fig. 5, where the resistors forming the axis have the value $2R$.

The network of fig. 5 is not used in practice; practical versions extend to one side of the z-axis only. Such networks are perfectly satisfactory if the axial resistors are given a value of $4R$ instead of $2R$. The validity of this can be confirmed by reasoning as follows. Imagine the $2R$ resistance along the z-axis in fig. 5 to have been replaced by two $4R$ resistances in parallel, as in *fig. 6a*. On account of the rotational symmetry of the system, no current flows from the upper portion to the portion under the z-axis. The lower portion can therefore be omitted (fig. 6*b*) without making any difference to the upper portion [12]).

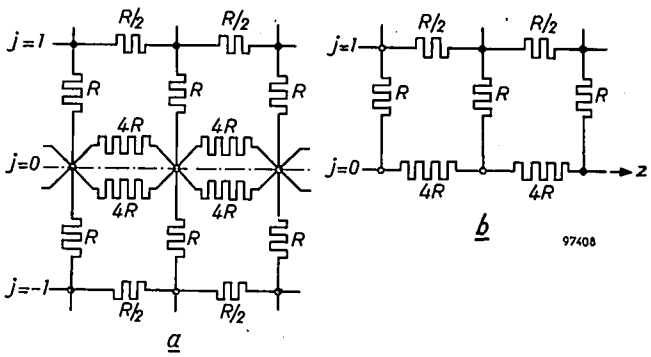


Fig. 6. Network (*a*) is equivalent to that in fig. 5, the $2R$ resistors along the z-axis having been replaced by two parallel $4R$ resistors. On account of symmetry, no current flows from the part of the network below the z-axis to the part above it. The lower half can therefore be removed without affecting the upper half, as in (*b*).

Design and use of the resistance network

Resistance networks suitable both for two-dimensional problems and for three-dimensional problems with rotational symmetry have been constructed in several Philips Laboratories. The two designs are identical apart from the values of the resistors and we shall therefore confine ourselves to describing the network for solving rotationally-symmetric three-dimensional problems. This network is constructed according to the arrangement shown in fig. 6*b*. It extends over 50 meshes in the z-direction and over 25 in the r-direction, and is accordingly composed of $(51 \times 25) + (26 \times 50) = 2575$ resistors in all, which are mounted on the back of a sheet of insulating material (*fig. 7*). The junctions have silver-plated contact pins that pass through to the front of the panel.

In order to determine the potential distribution in some electrode assembly (for example), the system is simulated on the resistance network by linking the junctions corresponding to the electrode outlines with copper wire. In principle it would be possible to apply voltages across the simulated electrodes in the manner described above (p. 13, first column). This is not necessary, however: by the

[11]) Along the axis, $\partial\varphi/\partial r$ and the other odd-order derivatives with respect to r are all zero in consequence of the symmetry of the system. Hence the middle term of differential operator (13) becomes 0/0; similar indeterminacies occur in the correction terms of (14). If the terms of the finite-difference operator (15) (with u replaced by φ) are arranged in a slightly different manner, the expression $(a/2r)(\varphi_Q - \varphi_P)$ appears which likewise has the value 0/0.

[12]) The network of fig. 6*b* can be arrived at directly by writing (19) in the form:

$$(\mathrm{L}w)_O = \frac{1}{a^2}\{(w_P - w_O) + (w_Q - w_O) + 4(w_S - w_O)\} .$$

On comparing (11) with the above, we obtain the condition

$$\frac{1}{R_1} : \frac{1}{R_2} : \frac{1}{R_3} : \frac{1}{R_4} = 1 : 1 : 0 : 4 ,$$

which is satisfied if $R_1 = R_2 = 4R$, $R_3 = \infty$ and $R_4 = R$.

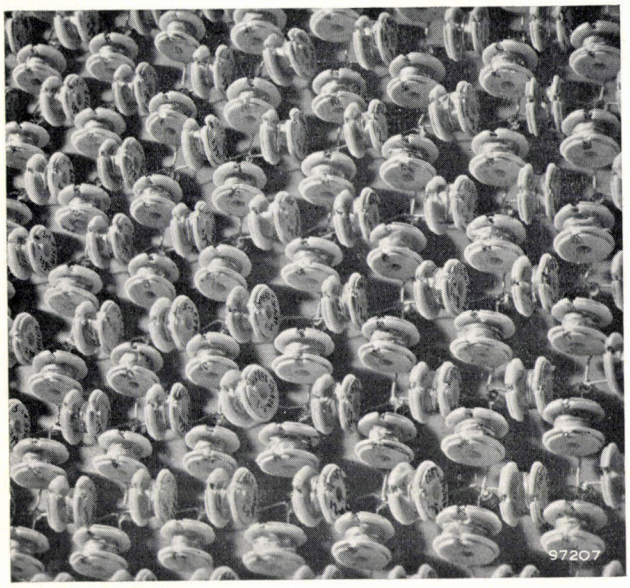


Fig. 7. Some of the 2575 resistors composing the network designed for three-dimensional problems with rotational symmetry.

following simple procedure the required potential distribution can be found more conveniently. One of the electrodes, G for example (see *fig. 8*), is connected to terminal B of potentiometer AB, and all the other electrodes (K and A_1 in fig. 8) are connected to A, the other potentiometer terminal. A and B are connected up to an accumulator. To measure the potential at, say, the junction P, the latter is connected to the slide contact of the potentiometer via a null indicating instrument. Once the slide has been brought to a position where the null indicator shows a null reading, the potentiometer setting indicates the potential difference between P and A (or B) as a proportion of the potential difference between B and A. The value thus found is therefore the potential of P when electrode G has unit potential and all the other electrodes have zero potential. By repeating the measurements for the other junctions one obtains the potential distribution under the above-mentioned circumstances. One of the other electrodes, A_1 for example, is now given an effective potential of unity and the others are held at zero; the potential distribution is then measured again. The potential distribution for any given combination of electrode potentials is then found simply by combining these results linearly (superposition). Generalizing, if there are n electrodes instead of three, the measurements have to be repeated $n-1$ times in order to find, by linear combination of the results, a solution for any given set of electrode potentials. The convenience of this method lies in the fact that no adjustment or measurement of voltage is necessary.

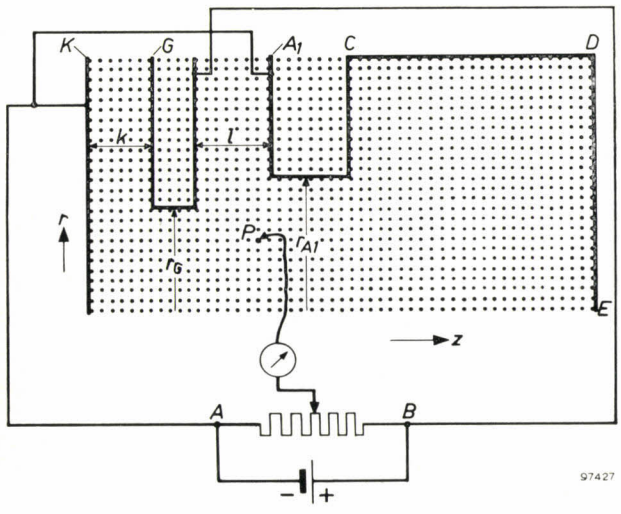


Fig. 8. Bridge circuit for measuring the potentials of junctions in the resistance network. K, G and A_1 are electrode models. The configuration is the same as in fig. 9. AB is a 1000 Ω potentiometer with a setting error of about 0.01 Ω. The null indicator is an electronic D.C. millivoltmeter with an internal resistance of 0.6 MΩ.

In view of the accuracy required, all the resistors were wound from manganin wire, to tolerances of $\pm 0.2\%$. Liebmann [7]) has pointed out that the average error arising in the measurement of potential, due to inexact resistance values, is much smaller than the errors in the resistances themselves, being from a tenth to a hundredth thereof. This is a consequence of the statistical properties of the network, whereby errors are levelled out. The temperature coefficient of manganin is so small, the voltage employed (usually about 2 V) is so low and the physical dimensions of the resistors are so large that there is no fear of errors due to heating-up of the resistors.

The upper limit to the (in principle, arbitrary) value of R (see fig. 6*b*) is fixed by the requirement that the highest value in the network, which is $4R$, shall not be an unreasonably high one for wire-wound resistors. On the other hand the smallest resistors must not have too low a value, otherwise the current through them would be large enough to set up appreciable potential differences in the copper wires representing the electrode outlines. In the present networks, R has the value 3600 Ω. The extreme resistance values are therefore $4R = 14\,400$ Ω and $R/50 = 72$ Ω.

A valve voltmeter type GM 6010 [13]) serves as the null indicator. It is a D.C. millivoltmeter combining

[13]) A. L. Biermasz and A. J. Michels, An electronic D.C. millivoltmeter, Philips tech. Rev. **16**, 117-122, 1954/55.

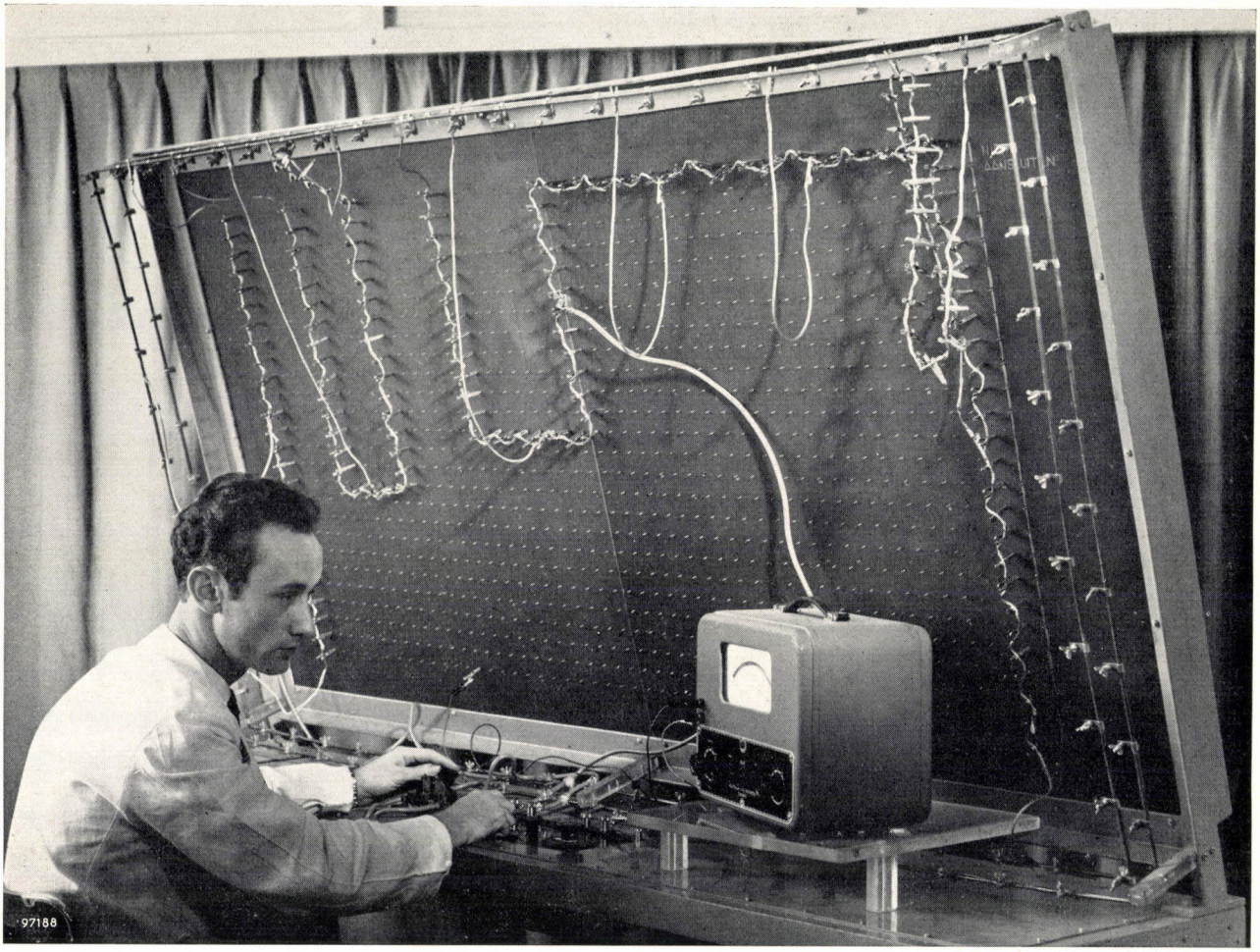


Fig. 9. The resistance network is mounted on the back of a large board of insulating material. The silver-plated contacts that are visible on the front of the panel are the junctions of the network. The electrode configuration seen on the board corresponds to a problem dealt with in this article by way of example. Lengths of 2 mm copper wire, representing electrode outlines, are attached to the appropriate contacts with metal clips. Electrical potentials are applied to these electrode models via heavy copper strips lying along the edges of the network. In order to prevent differences of potential arising along the electrode outlines, each is linked to its copper strip by more than one wire. In the photograph the user has his right hand on one of the knobs of the decade potentiometer; with his left hand he is pressing a key that enables him to check that the needle of the instrument he is watching is giving a null reading.

great sensitivity (readings down to 2 μV can be obtained) with a high internal resistance (0.6 MΩ). The "null current" is therefore less than about 3×10^{-12} A, which is so small that it makes no perceptible difference to the potential distribution. If it was other than very small it could give rise to appreciable errors, particularly in measurements on the axis of symmetry, where the highest-valued resistors lie.

The potentiometer must be very accurate, as its errors show up unchanged in the results. The potentiometer employed had an average accuracy of 1 in 10^5.

A photograph of the resistance network appears in *fig. 9*.

Fields of infinite extent

The region throughout which the required function is present is by no means always enclosed by a boundary such as s_3 in fig. 2. Often it extends to infinity, φ approaching a constant value φ_∞. One can proceed as follows in such cases. First of all, the electrodes are plotted on the network on a scale so small that the junctions around the edge correspond to grid points where φ is already fairly close to φ_∞. All the junctions around the edge are joined up and given a potential corresponding to φ_∞. The simulated electrodes are given potentials in proportion to those of the actual system and an equipotential curve, just enclosing the area within which φ has to be found, is then determined by measurement. Subsequently the scale is increased in such a way that it is still just possible to accommodate the equipotential curve thus found — whose potential is now known — on the resistance network; from then on this curve is treated as an electrode. A similar method can

be employed when it is desired to investigate only part of the region where φ is present, and to plot it on a very large scale, so that one or more electrodes or parts thereof fall outside the resistance network.

The same measurement procedure can be followed as before, one of the electrodes being given a potential differing from that common to other electrodes, and the nett potential distribution being found by linear combination of the results of successive measurements under these conditions.

It is often possible in practice to employ less laborious methods for circumventing the limitations of the resistance network due to its finite dimensions. We shall return to this point when discussing one of its applications.

Example of an application: the determination of the cut-off voltage and maximum cathode loading of electron guns

Often potential distributions are determined to serve as basis for the calculation of electron paths. Here we shall deal with a different example, namely the measurement of potential distributions within tetrode electron guns for cathode-ray tubes as a step towards determining the cut-off voltage and the maximum cathode loading. These two quantities are of importance in the design of electron guns.

For our present purposes an electron gun can be stylized as a set of parallel flat electrodes of infinite extent. The simplest form of gun (the triode gun in *fig. 10*) comprises three electrodes, the cathode K, the grid G and the anode A. Grid and anode have a circular aperture. The centres of the apertures lie on an axis perpendicular to the cathode, and consequently the potential field exhibits rotational symmetry about this axis. The anode has a positive voltage V_A, of 15 kV, say, with respect to the

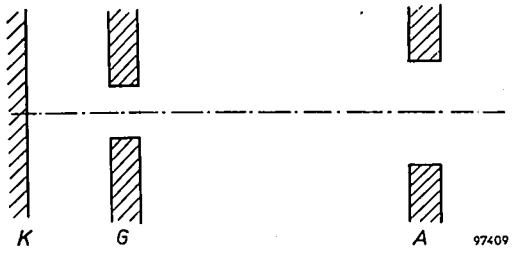


Fig. 10. Diagram to show the arrangement of electrodes in a triode gun. K cathode. G grid. A anode.

cathode. The grid voltage V_G (which is also measured with respect to the cathode) serves to modulate the beam current I issuing from the gun. For a given anode voltage V_A, V_G can be adjusted to a value such that the beam current is reduced to zero. This V_G value is always negative, of course; it is termed the cut-off voltage V_c. It is an important quantity, determining the maximum signal voltage that may be applied to the grid without its becoming positive (see *fig. 11*). Moreover, there is a simple relationship between V_c and the maximum current I_{max} that the gun can deliver under these conditions, viz. [14]):

$$I_{max} \approx 3\times10^{-6}\ V_c^{\frac{3}{2}} \text{ ampere} \quad . \quad . \quad . \quad (21)$$

(V_c in volts). It is therefore desirable that some method should be available for determining V_c from

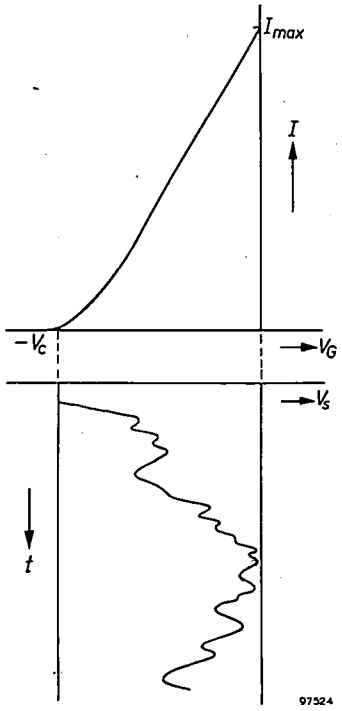


Fig. 11. The above curve showing the beam current I of a triode gun as a function of the grid voltage V_G, the anode voltage being constant, is a normal I_A-V_G characteristic for a triode. If excessive spot "blooming" is to be avoided, the signal voltage must not be allowed to push V_G above zero (otherwise grid current would start to flow). The "blacks" in the signal current correspond to $-V_c$, the cut-off voltage. Hence V_c also represents the maximum signal voltage.

the dimensions of the gun and from the voltage applied to the anode.

For this purpose we employ a graphical method, the graphs being derived from measurements performed with the resistance network. Use can be made of these graphs in the design of tetrode guns. A tetrode gun has a fourth electrode which is mounted close to the grid in the space between grid and anode (*fig. 12*). The fourth electrode is given a positive potential of about 300 V and is known as the "first anode", the original anode being called "final anode" in order to distinguish between the two. In practical cases the voltage on the final anode has but little influence on the cut-off voltage and the beam current, and consequently we only

[14]) See for example M. Ploke, Elementare Theorie der Elektronenstrahlerzeugung mit Triodensystemen, Z. angew. Phys. 3, 441-449, 1951 and 4, 1-12, 1952.

need consider the triode portion K-G-A_1. Even so, the system possesses properties different to those of a normal triode gun. In the latter, grid and anode are comparatively far apart. Consequently the field between these electrodes is more or less uniform and

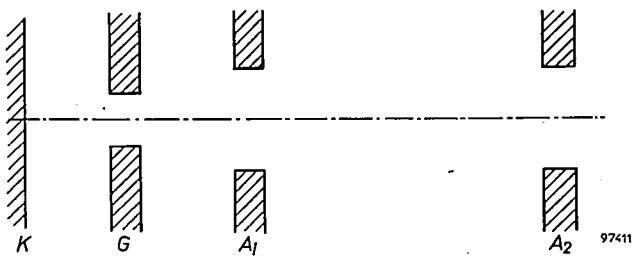


Fig. 12. In a tetrode gun the first anode A_1, which is given a potential of about 300 V, is placed close to the grid G. The final anode A_2 carries a high tension of about 16 kV.

its strength is given by the potential difference between grid and anode divided by the distance separating those electrodes. Hence this ratio determines the cut-off voltage, the diameter of the anode aperture playing no part at all. In a tetrode gun, on the other hand, the clearance between grid and first anode is small and the field between them is anything but uniform. Consequently the potential difference and distance between these electrodes each exercise a separate effect, and the size of the aperture in the first anode also has an influence on the cut-off voltage.

Let us take the following for the potential distribution along the axis (assuming $\varphi(0) \doteq 0$):

$$\varphi(z) = V_{A1} f(z) + V_G h(z) . \quad . \quad . \quad . \quad (22)$$

$f(z)$ represents the variations in potential along the axis when $V_{A1} = 1$ and $V_G = 0$, and $h(z)$ represents the corresponding distribution when $V_G = 1$ and $V_{A1} = 0$ (*fig. 13*). The beam current will be cut off when the current density in the centre of the cathode (where current density, as a function of position on the cathode, always exhibits a maximum) has become zero. Let us assume that the electrons have no initial velocity on quitting the cathode; if that is so, the beam current will be cut off the moment that the potential gradient at the centre of the cathode becomes zero, for the field strength $E(0)$ will then be zero at that point. Since E along the axis is $-\partial\varphi/\partial z$, we can find the field strength at the cathode centre by differentiating (22) with respect to z and putting $z = 0$:

$$E(0) = -\{V_{A1} f'(0) + V_G h'(0)\} . \quad . \quad (23)$$

$E(0)$ becomes zero when $V_G = -\{f'(0)/h'(0)\} V_{A1}$; hence the cut-off voltage is

$$V_c = \frac{f'(0)}{h'(0)} V_{A1} = D V_{A1} . \quad . \quad . \quad . \quad (24)$$

The quantity $D = f'(0)/h'(0)$ is known as the "penetration coefficient" or "Durchgriff".

With the aid of the resistance network, curves $f(z)$ and $h(z)$ were determined for many different combinations of electrode dimensions and separations, D being derived from the slopes of these curves at the cathode. It was found that their slopes at this point hardly alter in consequence of a change in the axial thickness of the first anode; hence D is virtually independent of that dimension. The discovery was welcome, because it meant one parameter less to be considered. Our investigations were limited to the case where the openings in grid and first anode have the same diameter. Since it is only the ratios between electrode dimensions that matter, the parameters are finally reduced to three. The overall results of the measurements are displayed in the set of graphs appearing in *fig. 14*. From these graphs one can determine the penetration coefficient of a gun whose dimensions are known; V_c, the cut-off voltage, can then be calculated with the aid of (24).

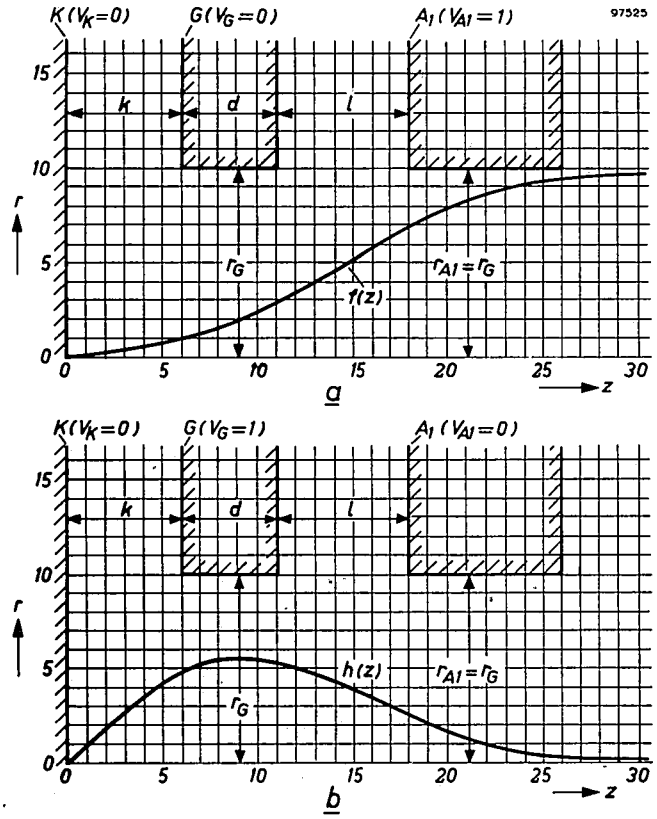


Fig. 13. *a*) Potential distribution $f(z)$ along the central axis in the triode portion of a tetrode gun when $V_G = 0$ and $V_{A1} = 1$. *b*) Potential distribution $h(z)$ along the same axis when $V_G = 1$ and $V_{A1} = 0$.

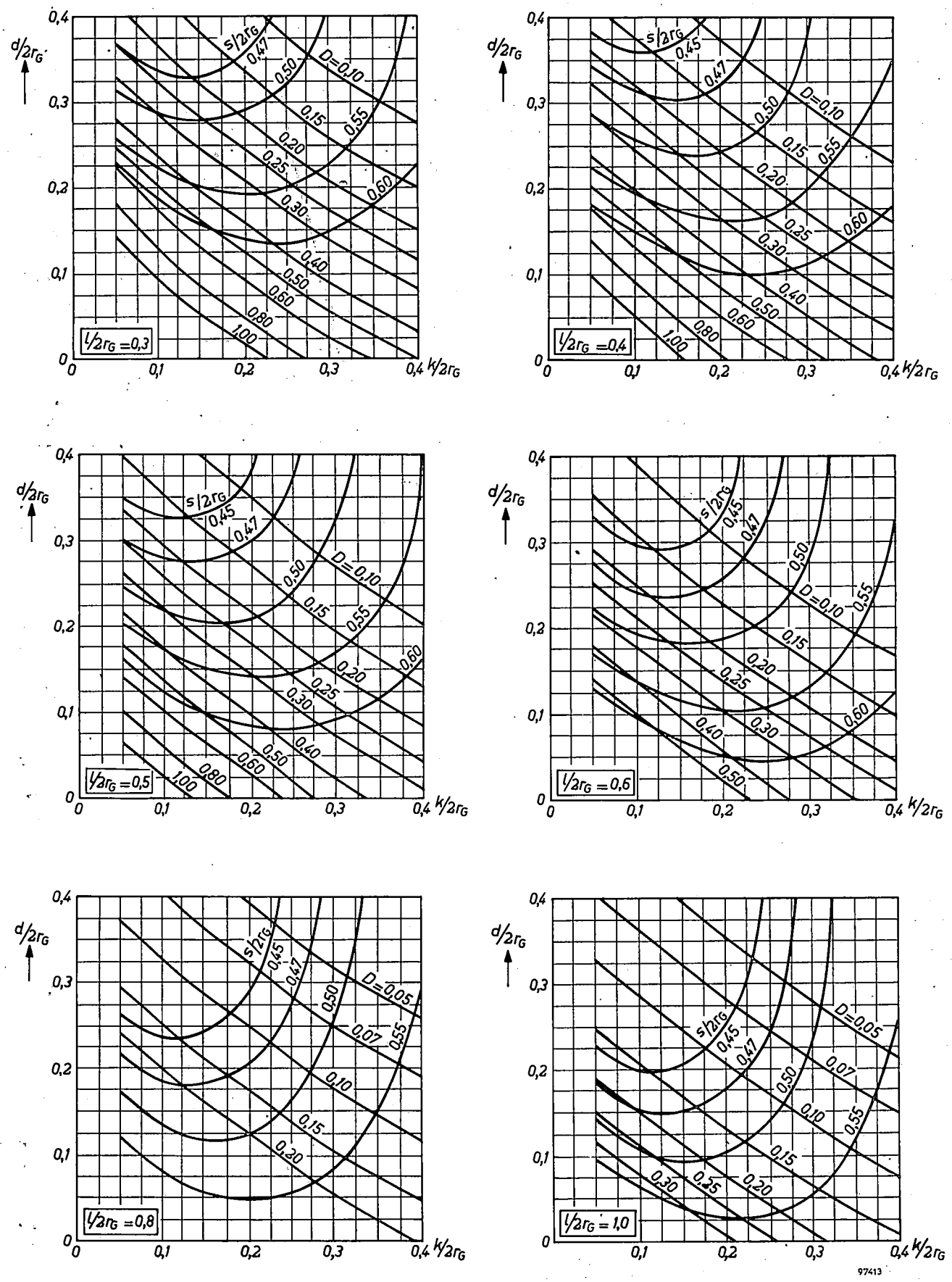


Fig. 14. Graphs allowing D, the "penetration coefficient", and s, the "equivalent anode distance", to be determined for an electron gun of given dimensions. For the meaning of the letters see fig. 13.

A second important quantity provided by the same measurements is the current density in the centre of the cathode (the maximum cathode loading). It is impermissible for the current density to exceed a certain value at this point, and this naturally constitutes a limit in design. Using a simplified theory, Ploke [14]) has derived the following for the current density at the cathode centre:

$$j(0) \approx 4.8\times10^{-3}\,(I\,V_c)^{\frac{3}{5}} \times s^{-2}\ \mathrm{A/mm^2}, \quad . \quad (25)$$

in which I is measured in amperes and V_c in volts, and where

$$s = \frac{1}{h'(0)}\ \mathrm{mm}.$$

The parameter s is sometimes called the "equivalent anode distance". The reason for the name is that if a potential difference of $V_c + V_G$ exists between cathode and anode in a plane parallel diode ($V_G < 0$), the field strength in the intervening space becomes equal to $E(0)$ when the anode-to-cathode distance is s. This can be deduced from (23) and (24).

In order to calculate D we had to determine the slope $h'(0)$ at the cathode; we can now use it again to calculate s. Curves from which s can be read off for a gun of known dimensions also appear in fig. 14. Knowing s, we can work out from (25) the maximum cathode loading for any value of beam current.

Numerical example

Suppose that we want to determine the cut-off voltage and the maximum cathode loading for a tetrode gun with the dimensions $k = 0.15$ mm, $d = 0.15$ mm, $2r_G = 2r_{A_1} = 0.75$ mm and $l = 0.35$ mm (see fig. 13), and with an accelerator anode potential of $V_{A_1} = 300$ V.

By calculation, $k/2r_G = 0.20$, $d/2r_G = 0.20$ and $l/2r_G = 0.466$. The values of D and $s/2r_G$ appropriate to these values of $k/2r_G$ and $d/2r_G$ are now determined from the two graphs in fig. 14 for $l/2r_G = 0.4$ and 0.5. We then find by linear interpolation that, for $l/2r_G = 0.466$, D is 0.23 and $s/2r_G$ is 0.51, from which it follows that s is 0.39 mm. Formula (24) gives $V_c = 0.23\times300 = 69$ V for the cut-off voltage and formula (25) gives $j(0) = 0.40\ I^{3/5}$ for the maximum cathode loading. If for example the gun provides a beam current of 100 μA, so that $I = 10^{-4}$ A, the maximum cathode loading will be $j(0) = 1.6\times10^{-3}$ A/mm².

The limited dimensions of the resistance network

Figs. 8 and 9 are related to the example just worked out; they represent, however, the more general case in which the grid and first anode apertures are unequal. We may now make some observations concerning the measures taken to allow for the fact that the resistance network is not infinitely large in relation to the electrode models plotted upon it. It will be seen from these two figures that the outline of the first anode has been extended along two edges of the network, the junctions on CD and DE having been wired together. As already stated, the results of the measurements are much the same whether the final anode is present or not, and we can accordingly leave it off the network. In these circumstances any point beyond the first anode and at some distance from its aperture will have a potential equal to that of the first anode itself. This will certainly be more or less true of points in space corresponding to the junctions along CD and DE. There is therefore no objection to giving these junctions the said potential. If they were not joined up, one would be completely in the dark about possible errors arising because the portion of the network to the right of the first anode is, as it were, left floating. In the event, the junctions along CD and DE have been short-circuited; and while it is true that the outline thus traced no longer conforms to that of the first anode under investigation, one does at least know that the discrepancy will not give rise to any serious error.

Our second observation concerns the interelectrode spaces between K and G and between G and A_1. The network only extends over a comparatively short distance in the radial direction; what sort of error does this give rise to? The results of measurements of potential in the inter-electrode spaces along the edge of the network (i.e. for maximum r) are reassuring, for the variation in the z-direction proves to be linear within the required limits of accuracy. This means that the height of the network (its extent in the r-direction) is adequate.

Implications of the finite mesh width

Obviously, the mesh with a of the theoretical grid cannot be reduced indefinitely, as this would lead to ever larger models on the resistance network. Hence the network will only provide approximate solutions to the problems worked out on it. In general, it is impossible on the basis of purely theoretical reasoning to estimate this fundamental error with any degree of accuracy. To get some idea of this accuracy, the network can be used for working out a problem whose exact solution is known, the potential distribution found with the network being compared with the known distribution. Alternatively, measurements can be performed for ever smaller mesh widths, a better approximation to the correct solution then being found by extrapolation [15]) [16]).

Cylindrical capacitor

Here we shall give some results of investigations relating to a cylindrical capacitor [17]). For such a capacitor the potential distribution can be worked out exactly. Further, it is possible by calculation to find the solution that would be obtained with an

---

[15]) L. F. Richardson, How to solve differential equations approximately by arithmetic, Math. Gazette **12**, 415-421, 1924/25.

[16]) R. Culver, The use of extrapolation techniques with electrical network analogue solutions, Brit. J. appl. Phys. **3**, 376-378, 1952.

[17]) J. C. Francken, Electron optics of the image iconoscope, thesis Delft, 1953, p. 36 et seq.

"ideal" resistance network. The error due to the finite mesh width can then be determined by comparing the latter solution with the exact one. Other errors, like those due to inaccuracies in resistor values or in the adjustment of the bridge, are thus excluded in this comparison.

If the ratio of the radii of the inner and outer electrodes is 1 : 10 and if the mesh width is made equal to the inner radius, the error is found to be, at worst, —0.7% of the voltage across the electrodes (the worst error arises at the grid points closest to the inner electrode). Near the outer electrode the error is only about —0.07%. If the mesh width is halved, the errors become about —0.2% and about —0.02% respectively. By correction of results by extrapolation the errors can be brought down to about —0.03% and about —0.001% respectively.

The problem of the cylindrical capacitor was also worked out on the actual resistance network, with the two mesh widths mentioned above. After correction by extrapolation, the results for junctions near the inner and outer electrodes differed from the calculated true values by —0.035% and —0.008% respectively. Differences between the experimental values and those calculated from the "ideal" network are due to the other errors referred to above, and have nothing to do with finite mesh width.

An electrostatic lens

Extensive investigations were also made into the part played by the finite mesh width when the problem is to determine the potential distribution along the axis of an electrostatic lens as in *fig. 15a*.

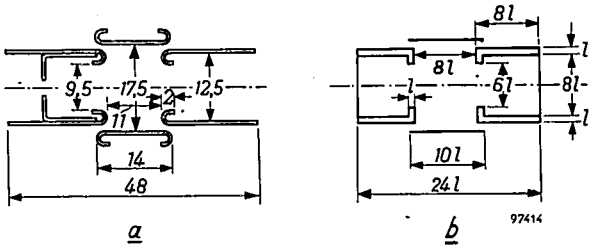


Fig. 15. *a*) Lens of a type used in electrostatically focussed picture-tubes. The electrodes to right and left have the highest potential in the tube; the middle one is at cathode potential. *b*) Shape of the same lens when stylized for the purpose of investigating the effect of the finite mesh width on resistance-network measurements of the potential distribution along the lens axis. The electrode dimension l has been selected as the characteristic length to fix the mesh number, i.e. the scale on which the lens is to be modelled on the network (mesh number $n = l/a$, a being the mesh width of the grid; see fig. 2).

This type of lens is used in television picture-tubes. The two narrower electrodes have the highest potential present in the tube; the middle electrode is at cathode potential. For the purpose of the investigation the lens was stylized, being given the symmetrical shape indicated in fig. 15*b*. In the stylized lens the ends of the outside electrodes are closed by conducting plates.

The electrodes are "capped" in this way for the same reason that led us to introduce further connecting wires into the model of the first anode dealt with above. Here as before, we have to consider the inter-electrode gaps. It is possible, amongst other things, to short-circuit the entire upper edge of the network and to give it zero or unit potential. It proves that this has no appreciable effects on the results. It may be concluded that the limited size of the network in relation to the spaces between the electrodes, does not prejudice the reliability of the measurements.

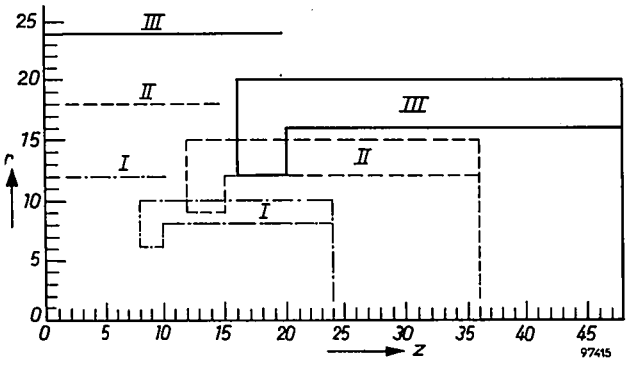


Fig. 16. The lens of fig. 15*b* plotted on the resistance network. Symmetry makes it unnecessary to model more than a quarter of the lens assembly. Measurements were carried out on three electrode models (one at a time, of course) plotted with mesh numbers of 2, 3 and 4, and marked *I*, *II* and *III* respectively in the diagram.

The left-hand edge of the resistance network is made a mirror plane of symmetry by doubling the values of the resistors forming this edge (cf. the

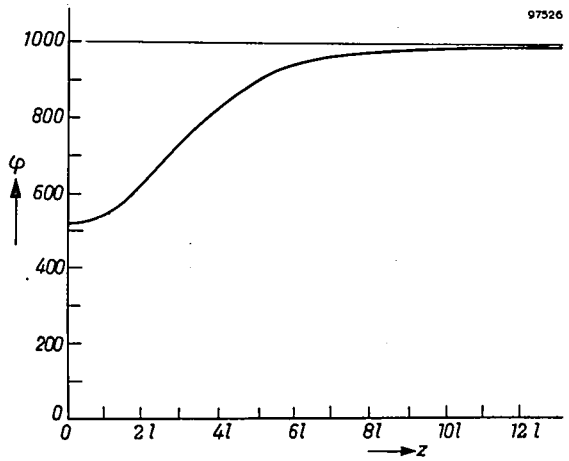


Fig. 17. Variation in potential along the axis of the lens in fig. 15*b*, as determined by means of a resistance network. The scale of this graph is too small to allow any distinction to be made between the slightly different results obtained from measurements on the three different models.

mirror plane along the z-axis in fig. 6*b*). In consequence, only a quarter of the lens had to be simulated on the network. The proportions of the stylized electrodes were so chosen that the lens could be plotted with "mesh numbers" that were in the proportions of 2 : 3 : 4. The "mesh number" is the ratio n between l, one of the dimensions of the lens, and a, the mesh width of the grid. Any dimension of the lens can be chosen for this purpose provided that it is consistently adhered to. Our choice is indicated in fig. 15*b*. The greater the mesh number, the finer is the grid and the bigger is the electrode model on the network (*fig. 16*). The potential values measured along the lens axis are plotted in *fig. 17*. The curve as drawn here is not thin enough to reveal divergencies arising from the use of the three mesh numbers.

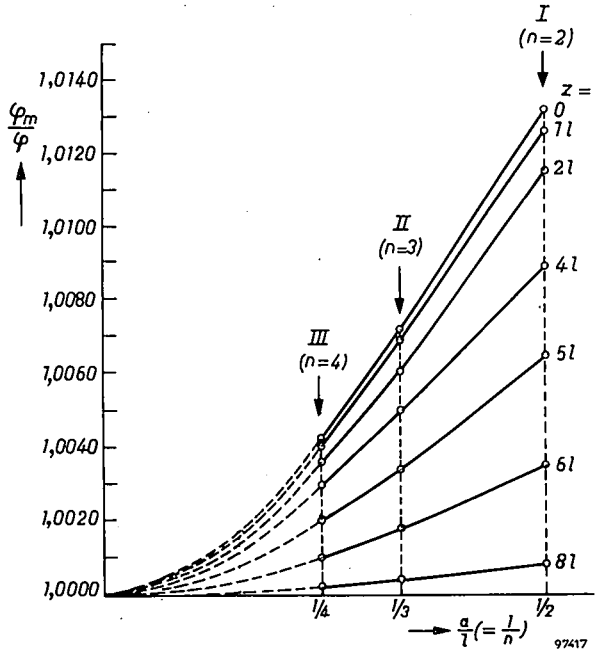


Fig. 18. The relative error (the measured value φ_m divided by the "correct" value φ) in the potential distribution curve of fig. 17, plotted as a function of the mesh width a. The quantity a/l along the abscissa is the reciprocal $(1/n)$ of the mesh number. The seven curves are appropriate to seven points along the z-axis. Each has been drawn through three points determined by resistance-network measurements on three models (*I*, *II* and *III* in fig. 16). Even when the coarsest grid is used the errors remain small.

The three values found for each point along the axis were used to work out, by Culver's method of extrapolation [16]), a value regarded as correct. The three measured values of each set were divided by the "correct" value and the results plotted in *fig. 18*; the three points of each set are joined by a smooth curve. Sets of values appropriate to several points on the z-axis are given. The figure clearly reveals how slight an effect is exercised by the finite mesh width on the results of measurements with the resistance network.

Summary. The resistance network has won a place beside the electrolytic tank as an aid to the solution of Laplace's equation for given boundary conditions. Two kinds of network have practical importance. The first is useful for solving two-dimensional problems, the second for solving three-dimensional problems where rotational symmetry exists. In either case Laplace's equation reduces to a differential equation involving two independent variables only. The network is built up of resistors, four of which meet at each junction. The junctions correspond to the grid points of a hypothetical square grid which is imagined to have been set up in the field space. Junctions corresponding to boundaries are given potentials proportional to those values which the boundaries are known to have. Provided the resistors composing the network have the right values, the remaining junctions will then acquire potentials that are approximately proportional to the required potential function. Discrepancies from the actual values decrease as the mesh width of the grid is reduced to zero. Conditions which must be satisfied by the resistances composing the network, in both the two-dimensional and in the rotationally-symmetric three-dimensional case, are worked out in the course of the article. An example of the employment of a network for rotationally-symmetric three-dimensional problems is given in which curves are derived which allow the cut-off voltage and maximum cathode loading of tetrode electron guns to be determined. Finally, examples are given to show that the errors due to the finite mesh width are very small.

VECTOR-ELECTROCARDIOGRAPHY

by G. C. E. BURGER *) and G. KLEIN.

621.3.012.1:612.17:612.172.4

Developments in electronic engineering enable the physician to apply increasingly refined physical methods to problems of diagnosis. One of these methods, which has made great strides in the last 20 years, is the investigation by means of the vector-electrocardiograph of the electrical phenomena accompanying the contraction of the heart. The article below describes one of the vector-electrocardiographs built in the Philips Research Laboratory at Eindhoven and in use at the Philips Health Centre. In the introduction the physical principles of vector-electrocardiography are touched on and a brief discussion is devoted to the nature of the electrical phenomena concerned and the relation between vector-electrocardiography and conventional electrocardiography from which it evolved.

Introduction

A valuable aid in cardiology is the study of the electrical phenomena that occur during the contraction and subsequent relaxation of the heart muscle. These electrical phenomena are related to the manner in which the stimulus giving rise to the contraction is propagated over the heart muscle. Changes in these electrical phenomena enable the cardiologist to learn something about the cardiac disorders responsible for the changes, such as the presence of an inactive part (infarction) or an enlargement (cardiac hypertrophy) of the heart muscle. Purely mechanical disorders (openings in the septum, valvular deficiencies) are not primarily included in this category, although they are not infrequently the cause of other disorders which in turn do manifest themselves by an electrical phenomenon.

Electrical phenomena accompanying muscular contraction

The electrical phenomena referred to are brought about roughly as follows.

In the resting state the wall of every muscle fibre (a membrane) constitutes an electric double layer; the potential of the interior of the fibre is about 60 mV lower than that of the environment (the "medium"). This phenomenon is termed *polarization*. At the position where a stimulus is applied to the fibre, this potential difference — known as the *membrane potential* or *resting potential* — decays very rapidly (in approximately 1 millisecond) and even reverses its polarity to the peak value of about 35 mV; this is known as *depolarization*. As a result of this effect an electric current, which varies from place to place throughout the whole volume of the medium, flows from the intact fibre wall to the stimulated part. This current becomes manifest at the boundary of the medium, in this case the surface of the body, in potential differences between the various points of the surface (*fig. 1*).

The current in the medium can be formally represented as being generated by a galvanic cell whose terminals are at a small distance apart on the centre-line of the fibre (dipole). The positive pole

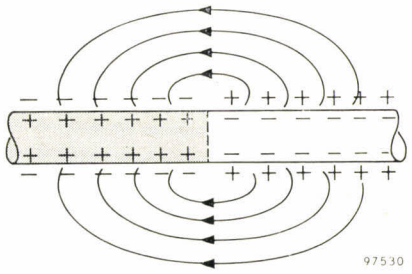


Fig. 1. Schematic representation of a partially stimulated muscle fibre. The non-stimulated part has a potential lower than that of the medium; in the stimulated part (shaded) the potential difference has changed its polarity. This causes a current to flow in the medium.

represents the intact part, the negative the stimulated part. The electric field is now determined by a quantity depending both on the potential difference between the terminals (the electromotive force) and their distance apart, and on the direction of the centre-line referred to; in other words, it is determined by a *vector quantity*.

The current flow implies that the medium surrounding the intact wall, particularly near the edge of the stimulated zone, loses charge, resulting in a drop in the potential difference between the two sides of the wall. This in turn acts as a stimulus, causing the membrane potential here also to change as described above; in other words, the boundary between stimulated and non-stimulated zones propagates itself along the length of the fibre. In regard to the fibres of the heart muscle the velocity of propagation is of the order of magnitude of half a metre per

*) Professor of Public and Industrial Hygiene, University of Amsterdam. Until his recent retirement Prof. Burger was the Director of Medical Services of N.V. Philips.

second. It should be noted that the transition region between the stimulated and non-stimulated part of these fibres is extremely short, and is in any case very small compared with the dimensions of the heart.

Immediately after a certain part of a fibre has entered into the stimulated state, the membrane potential begins to return to the resting value — the repolarization process — which it reaches in about 0.2 second. This relatively slow recovery means that the stimulated part of a fibre could have a length of about ten centimetres if the fibre were only long enough. In reality the length of the fibres is much less. This is one reason why the human heart can be in a completely depolarized state for a short time.

The electrical action of the boundary between a stimulated and a recovered region is analogous to that of the front edge of the stimulated region. The vector quantity associated with the repolarization, however, assumes an appreciable value only after some tenths of a second from the beginning of the stimulus.

An electric double layer with properties corresponding to those of muscle fibres is found, during life at least, in the walls of all biological structures that react to stimuli (nerves, muscles, glands, sensory receptors, etc.). The potential difference is a consequence of the permeability of the membrane, which differs widely for different kinds of ions. The membrane is completely impermeable to SO_4^{2-} ions and to the large organic ions contained in the fibre; to Na^+ it is only slightly permeable, whereas it is relatively highly permeable to K^+ and Cl^-. As a result the various kinds of ion inside and outside the fibre differ considerably in concentration. The membrane potential is virtually equal to the diffusion potential of the K^+ ions. The action of a stimulus causes the permeability to Na^+ ions to become suddenly about 500 times greater, which then far exceeds for a short time the permeability to K^+ ions. Consequently the membrane potential changes sign and becomes virtually equal in magnitude to the Na^+ diffusion potential. Na^+ ions then flow in and K^+ ions flow out.

When the stimulus ceases, the permeability returns to its original value. The K^+ ions diffuse back again through the wall, but the Na^+ ions are now "trapped" inside and are "pumped" out by a metabolic process known as the sodium pump mechanism.

Although the electrophysiological study of the nervous system has made great advances, particularly in the last ten years, it is still not yet known exactly how the Na^+ pump works, nor how it comes about that the application of a stimulus suddenly changes the permeability of the wall to Na^+ ions. The restoration to the resting state also awaits clarification [1]).

The heart vector

The electric field which the heart muscle as a whole gives rise to in the body is the sum of the contributions of each separate fibre. At a large distance from the source this electric field cannot be distinguished from that of a single fibre (dipole) the electromotive action of which is characterized by a vector equal to the sum of the vectors which define the action of each of the fibres belonging to the muscle. In close proximity to the heart this is no longer applicable; here the closest parts of the muscle make a greater contribution to the field at a certain point than the parts farther removed.

This vector sum, which thus defines the electromotive action of the whole heart muscle as manifested at a relatively large distance in the medium, is termed the *heart vector* (*fig. 2*). Because of the complicated way in which the heart muscle is depolarized and repolarized, this vector undergoes marked

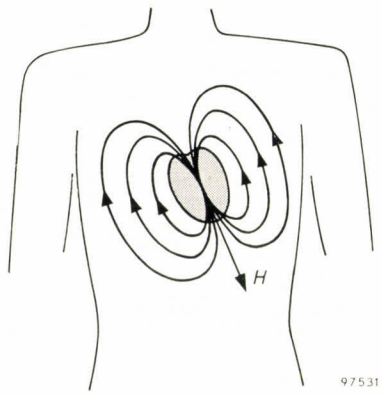


Fig. 2. The electrical action of the heart as manifested at a relatively large distance from the heart in potential differences between the various points of the medium can be described by a vector **H**. During the heart beat this vector varies considerably in magnitude and direction.

changes in magnitude and direction during the heart beat. If we represent the heart vector by an arrow one end of which is at a fixed point, the arrow will thus swing to and fro, while varying in length, around this fixed point during the heart beat. The point of the arrow describes in space an irregular curve, called the *vector-electrocardiogram* (Vcg).

Conventional and vector-electrocardiography

As mentioned, the current flowing through the body gives rise to potential differences between the various points on the surface of the body. These potential differences are of the order of magnitude of 1 mV. With a sensitive and rapidly indicating instrument of high internal resistance, these potential differences can be measured and their variation with time recorded. The result is a conventional electrocardiogram (Ecg). Its appearance varies, of

[1]) A review of the present state of knowledge in this field will be found in A. von Muralt, Neue Ergebnisse der Nervenphysiologie, Springer, Berlin 1958, and in J. C. Eccles, The physiology of nerve cells, Johns Hopkins Press, Baltimore 1957.

course, with the chosen combination of points between which the measurement is made.

Such a combination of two contact points is known as a *lead* or *derivation*. The number of possible leads is obviously unlimited, but for practical reasons the electrodes are applied to a number of standard positions, the oldest of which are the two arms and the left leg. Standard positions have also been selected on the back and chest; sometimes, too, an electrode is let down into the oesophagus. In a complete cardiological examination electrocardiograms are recorded as a rule from a large number of leads.

It should be noted that some leads are not obtained by connecting the instrument directly to two electrodes, but by connecting only one terminal directly to the electrode and the other via resistors of about 5000 Ω to three others placed on the two arms and the left leg. The potential on this terminal is then equal to the average of the potentials on the three electrodes to which it is connected; this does not imply, however, that this average is zero or that it has any other value not varying with time. The leads between electrodes on the arms and the left leg are known as the three *standard leads of Einthoven*. Since the current lines hardly penetrate into the extremities, it is of little consequence where exactly the electrode is placed, which makes it easy to work with these leads.

To obtain good recordings it is of course necessary that the potential differences applied to the instrument should be caused solely by the action of the heart. The patient should therefore be completely relaxed. The potential differences originating from other muscles are often the chief form of interference.

Form of conventional and vector-electrocardiograms in relation to the action of the heart

Before considering the manner in which a vectorcardiogram is recorded, it will be well to take a closer look at the action of the heart muscle and the way in which it manifests itself in electrocardiograms and vectorcardiograms (we shall frequently below use this abbreviated form in place of vector-*electro*cardiogram(graphy)).

As is well known, the heart consists of two parts, the left heart and the right heart, which are separated by a dividing wall or *septum*. Each part has two chambers, i.e. an *auricle* (or *atrium*) where the blood enters, and a *ventricle*, which forces the blood into the relevant arteries. The right heart receives the venous blood depleted in oxygen and pumps it into the lungs, and the left heart receives the oxygenated blood from the lungs and pumps it through the arteries to supply the organs and tissues of the body (*fig. 3a*). Both halves of the heart can in fact be regarded as two separate pumps connected in series (fig. 3*b*). The vascular systems between the two pumps are known respectively as the *lesser* (pulmonary) circulation and the *greater* (systemic) circulation.

The interauricular septum, i.e. the part of the dividing wall between the auricles, is a membrane; the outer wall of the auricles is a fairly thin muscle.

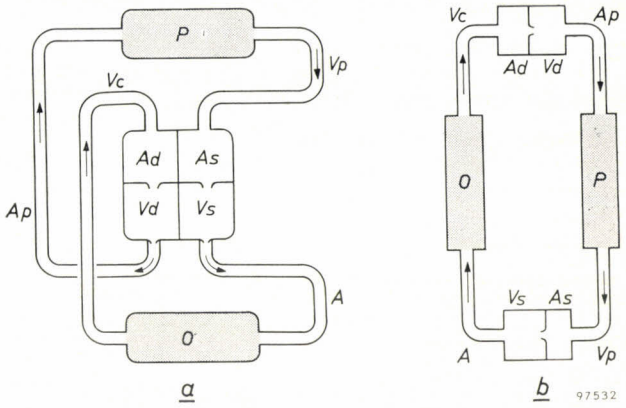


Fig. 3. *a*) Schematic representation of the human circulatory system. *As* left auricle, *Vs* left ventricle, *A* aorta, *O* organs, *Vc* vena cava, *Ad* right auricle, *Vd* right ventricle, *Ap* pulmonary artery, *P* lungs, *Vp* pulmonary vein. *b*) The same with the two halves of the heart drawn separately. These halves constitute simultaneously operating pumps connected in series.

The upper part of the septum between the ventricles is also a membrane, but somewhat lower it goes over into a muscular wall which is integral with the outer wall of the ventricles. This structure is known as the heart muscle or *myocardium*. The upper part is called the *basis*, the bottom part — i.e. the point of the heart — is the *apex* (*fig. 4*).

Depolarization takes place in the following way. The stimulus begins in the *Keith-Flack node* (or

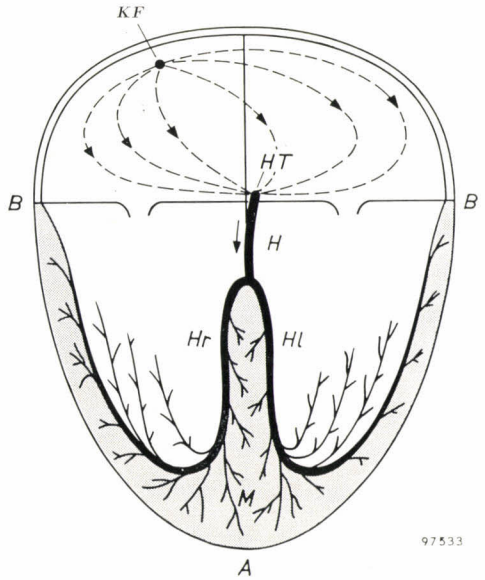


Fig. 4. Representation of stimulus conduction in the heart. *A* apex, *B-B* basis, *M* heart muscle (myocardium), *KF* Keith-Flack node (sino-auricular node), *HT* His-Tawara node (atrio-ventricular node), *H* unramified part of the bundle of His (atrio-ventricular bundle), *Hl* left branch of bundle, *Hr* right branch of bundle.

sino-auricular node) which is located in the wall of the right auricle. From this point the stimulus spreads out over the wall of both auricles. The auricles contract at virtually the same time. At a certain moment the stimulated zone reaches the *His-Tawara node,* which is situated in the septum near the basis of the myocardium (for which reason it is also called the *atrio-ventricular node*). The latter node is the source of a special system of stimulus conduction. This involves a bundle of fibres of non-contractile muscle tissue (known as the *bundle of His* or *atrio-ventricular bundle*) in which stimuli are propagated at about 1 metre per second. The bundle splits into two branches at the point where the septum ceases to be membranous and merges into the myocardium; one branch continues at the left of the septum, the other at the right. These branches in turn split up into large numbers of fibres which extend in the interior of the myocardium over the entire wall of both ventricles, and from which still finer branches extend into the myocardium itself. From these branches of the bundle of His the stimulus propagates itself in the fibres of the ventricular muscle. After some tenths of a second repolarization begins.

The three successive events, depolarization of the auricles, depolarization of the ventricles and repolarization of the ventricles, give rise in the electrocardiogram and the vectorcardiogram to three clearly distinguishable deflections. These are denoted respectively by *P*, *QRS* and *T*. *Fig. 5* shows a normal electrocardiogram obtained by recording the potential difference between the right and left arms. It can be seen that between the three electrical phenomena there are small time intervals during which the potential difference is zero. In the *vectorcardiogram* the three phenomena appear in the form of three loops, the starting and terminal points of which, for a normal heart, lie in the origin. The form of the *QRS* loop, and whether or not this loop returns to the origin before the *T* loop begins, are of great significance to the cardiologist. The form of the *P* loop is, as yet, of no clinical importance.

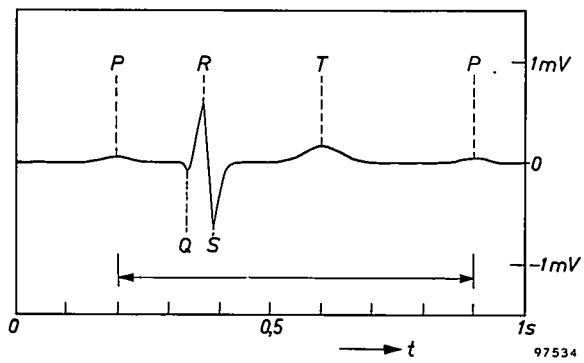


Fig. 5. Electrocardiogram of a normal heart as obtained when the electrodes are applied to the right and left arms. There are three distinct parts which correspond, respectively, to the depolarization of the auricles (*P*-spike), that of the myocardium (*QRS*-complex) and the repolarization of the myocardium (*T*-spike). The length of the arrow corresponds to the time elapsing between two heart beats.

Finally it should again be noted that the heart vector is the sum of a large number of individual vectors differing in direction and magnitude. The vectorcardiogram provides no information on these individual vectors as such. The relation between the direction of the heart vector and the stimulated state of the heart at a given moment should not be over-simplified. The direction of the vector can be simply correlated with the direction in which the change of state is propagated only during the very first beginning of myocardiac depolarization (start of the *QRS* loop) and during the very first beginning of repolarization.

Mathematical principles of vectorcardiography

In vectorcardiography the object is to try and find the form of the heart vector itself — the above-mentioned vectorcardiogram [2]) — rather than to record a number of separate electrocardiograms. Assuming that the electric field appearing at the surface of the body may be regarded as excited by a stationary point-source dipole, only three mutually independent leads are sufficient for this purpose.

Since the dimensions of the heart are not so very small compared with the body, the above assumption only roughly corresponds to reality at places on the surface of the body which are not too near to the heart. Further, just as in ordinary electrocardiography, the electrodes should obviously not be too close together. It is often found convenient to place three of the four electrodes on the extremities and the fourth somewhere on the trunk. We shall return to this subject presently.

The problem of how the magnitude and direction of the heart vector are related to the potential differences between the electrodes — a problem purely of mathematical physics — is solved in the following way [3]). It can be deduced from physical considerations that the relation between the three independent potential differences P_1, P_2 and P_3 and the magnitude of the three mutually perpendicular components X, Y and Z into which the heart vector

[2]) Research in this field was pioneered by H. Mann (Arch. int. Med. **25**, 283, 1920). The first vectorcardiograph was described by H. E. and W. Hollmann, Z. Kreislaufforschung **29**, 465, 1937.

[3]) H. C. Burger and J. B. van Milaan, Brit. Heart J. **8**, 157-160, 1946.

H can be resolved, is linear, and can thus be defined by the formulae:

$$P_1 = a_1X + b_1Y + c_1Z, \quad . . \quad (1a)$$

$$P_2 = a_2X + b_2Y + c_2Z, \quad . . \quad (1b)$$

$$P_3 = a_3X + b_3Y + c_3Z. \quad . . \quad (1c)$$

The coefficients a_1 to c_3 are constants which (in addition to being determined by the directions selected for the coordinate axes; see *fig. 6*) are

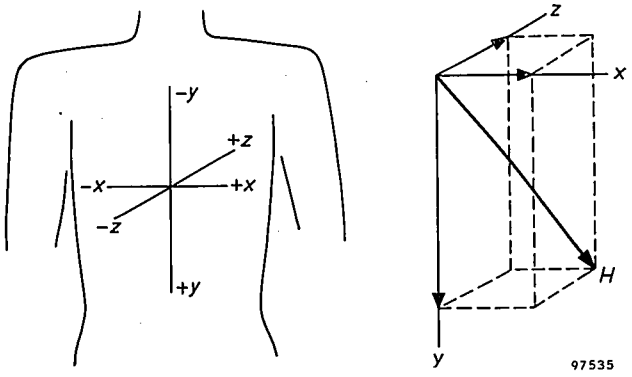


Fig. 6. Coordinate axes selected for vectorcardiography in accordance with the convention adopted by the American Heart Association. XY is the frontal plane, XZ the horizontal plane and YZ the sagittal plane.

determined by the shape of the various parts of the body (muscles, lungs, etc.) and their relative conductivity, by the position where the dipole is imagined to be located in the body, and finally by the position of the electrodes.

The values assumed by the coefficients a_1 to c_3 for various combinations of leads have been determined from measurements on a model of the human body [4]), complete with "lungs" and a "spinal column". The material chosen for the lungs was such that the ratio between their electrical conductivity and that of the rest of the trunk was approximately equal to that believed to exist in the body. The spinal column was made of a very poorly conductive material.

By solving equations (1) we find for the components of the heart vector:

$$X = \alpha_1P_1 + \alpha_2P_2 + \alpha_3P_3, \quad . . \quad (2a)$$

$$Y = \beta_1P_1 + \beta_2P_2 + \beta_3P_3, \quad . . \quad (2b)$$

$$Z = \gamma_1P_1 + \gamma_2P_2 + \gamma_3P_3. \quad . . \quad (2c)$$

Here again, the coefficients α_1 to γ_3 are constants, and can be calculated from a_1 to c_3. Since (2) has general validity, we must always find the same X, Y and Z when comparing measurements made with different leads, provided we use in the calculation those values for α_1 to γ_3 appropriate to the relevant leads. By carrying out comparative measurements of this kind on patients, we thus have a means of verifying the correctness of the values of a_1 to c_3 obtained on the model. In addition we can ascertain in how far the spatial extension of the heart (deviation from an ideal dipole) affects the results.

Recording a vectorcardiogram

In the vectorcardiograph discussed in this article the components of the heart vector are automatically computed from (2). After amplification, the potentials P_1, P_2 and P_3 are fed to a number of potentiometers to obtain the correct fraction of each of the signals. These fractions are then added together. The apparatus does this for each of the equations (2), so that three signals are obtained which are proportional to X, Y and Z.

The form of the earlier-mentioned curve, described in space by the point of the heart-vector arrow, is now found by successively applying combinations of two of the signals to the two pairs of deflection plates of a cathode-ray oscilloscope. If we select, for example, the signals which represent the X and Z components, the screen displays the projection of this curve in the horizontal plane. Analogously, X and Y yield the frontal and Y and Z the sagittal projection. By suppressing the electron beam at regular intervals (e.g. 1/200 sec), time marks are obtained by means of which the points corresponding to each of the various projections can be recognized. By interrupting the recording before the end of the heart beat, or by varying the line thickness between each two time marks, one can determine the direction in which the line is being described.

With some practice it is readily possible to construct from two of the three projections, using the time marks, a wire model corresponding to the three-dimensional curve. The third projection is thus not really necessary, although it is often easier to make the wire model when three projections are available.

Fig. 7 shows various projections and *fig. 8* the wire models derived from them. If the time marks are added to these models, we have an indication of the speed at which the point of the arrow covers the various parts of the loops.

Reverting to the end of the foregoing section, it may be mentioned here that investigations using a number of electrode configurations with known coefficients show that the agreement between the vectorcardiograms obtained with these on any patient is highly satisfactory in a large percentage of cases (70 to 80%) as far as the frontal projection is

[4]) H. C. Burger and J. B. van Milaan, Brit. Heart J. **9**, 154-160, 1947. See also Acta med. Scand. **114**, 584, 1943.

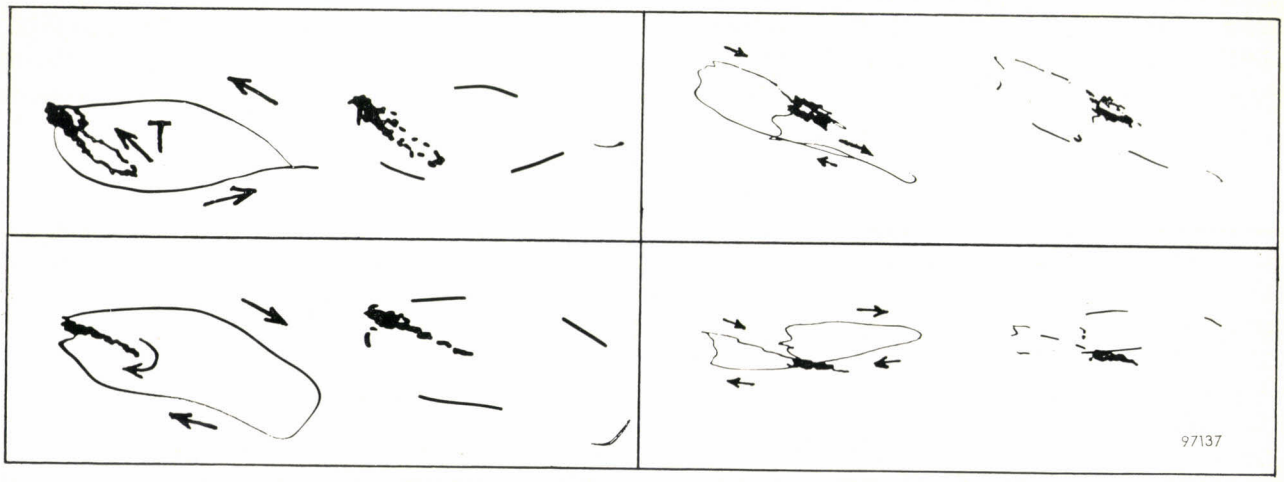


Fig. 7. Horizontal projection (above) and frontal projection (below) of the vector cardiogram of a normal heart (left column) and of a disordered heart (right column). For one of every pair of otherwise identical oscillograms the electron beam is interrupted 200 times per second to obtain time marks.

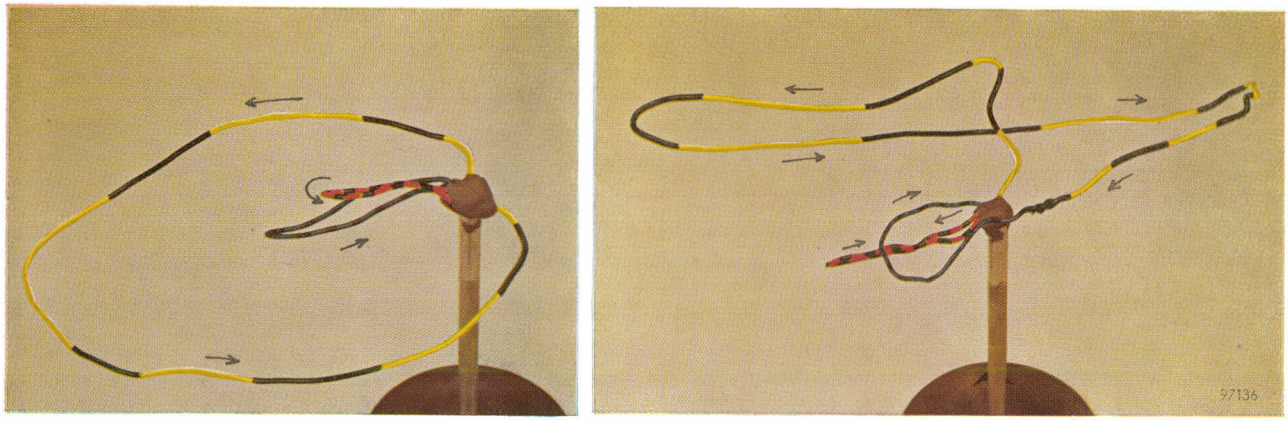


Fig. 8. Vectorcardiogram in the form of a wire model of the normal heart (left) and the disordered heart (right) derived from the two projections shown in fig. 7. The parts (loops) that correspond to the depolarization of the auricles (*P*; red), to the depolarization of the ventricles (*QRS*; yellow) and to the repolarization of the ventricles (*T*; blue) all begin and end at the origin in the case of the normal heart. With the disordered heart, this is not so as far as the *QRS* and *T* loops (yellow and blue) are concerned. The arrows indicate the direction in which the three loops are described. (The patient is viewed obliquely from above with his back to the observer.)

concerned. The agreement is not so good, however, between projections in which the *Z* component (antero-posterior) is involved. Further investigation has shown that the agreement cannot be improved by choosing other coefficients. This suggests that the position of the electrodes does not entirely comply with the physical conditions that must be satisfied if the heart vector (the vector sum of the component vectors) is really to be obtained. Since the electrodes obviously cannot be positioned farther from the heart than the dimensions of the trunk permit, systems employing more than four electrodes are now being studied. The use of more electrodes makes it possible to reduce the influence of the spatial configuration of the heart, i.e. to record the heart vector with less distortion [5]). The instrument about to be discussed was therefore designed for use with both a four-electrode and a five-electrode system. On the

[5]) A system using five electrodes was described by H. C. Burger in Annals New York Acad. Sciences **65**, 1076-1087, 1959, one with six by O. H. Schmidt and E. Simonson in Arch. int. Med. **96**, 574, 1955, and one with seven electrodes by E. Frank in Circulation **13**, 737, 1956.

In a theory put forward by D. Gabor and C. V. Nelson (J. appl. Phys. **25**, 413-416, 1954) it is shown that the fact that the heart is not small with respect to the dimensions of the chest does not in principle prevent the magnitude and direction of the heart vector (and even its position) from being determined with considerable accuracy. For this purpose, however, a very large (theoretically infinite) number of electrodes is needed. With a view to finding a reliable and at the same time practicable method, investigations are proceeding to ascertain how the number of electrodes can be reduced without unduly affecting precision.

other hand it should be noted that where there is poor agreement between the vectorcardiograms obtained with these two lead systems, the deviations are very seldom so serious as to cause significant clinical errors even if the physician were to rely solely on the vectorcardiogram. They do, however, call for some prudence when interpreting a vectorcardiogram.

The geometric relation between vector- and conventional electrocardiograms

The relation between the vectorcardiogram and conventional electrocardiograms obtained with three leads is given analytically by formula (1). If we realize that the potential difference P_1 in formula (1a), for example, is a *scalar* quantity, and that X, Y and Z are *vector* components, we can conclude that a_1, b_1 and c_1 must necessarily also be the components of a vector $\mathbf{p}_1$. P_1 is the *scalar product* of the vectors $\mathbf{p}_1$ and $\mathbf{H}$. Likewise P_2 is equal to the scalar product of $\mathbf{p}_2$ and $\mathbf{H}$, and so on. The vectors $\mathbf{p}_1$ etc. are called *lead vectors*.

On the other hand the magnitude of the scalar product of two vectors can be found by multiplying the magnitude of one of the vectors by the magnitude of the projection on to this vector of the other. Thus:

$$P_1 = p_1 H \cos \varphi, \quad \ldots \ldots \quad (3)$$

where φ is the angle between $\mathbf{p}_1$ and $\mathbf{H}$ and p_1 and H their modulii. If we now construct a tetrahedron the direction and length of three of whose edges correspond to those of the lead vectors $\mathbf{p}_1$, $\mathbf{p}_2$ and $\mathbf{p}_3$, we can deduce from the above that we shall find P_1 by projecting the heart vector $\mathbf{H}$ on the edge $\mathbf{p}_1$ (a_1, b_1, c_1) and by multiplying the length of the projection (which is $H \cos \varphi$) by the length of this edge, and so on.

Of course we can construct such a tetrahedron for any four electrode positions. We select three leads in such a way that they have one electrode in common, and we construct the tetrahedron such that the lead vectors meet in one corner. By keeping one point fixed when changing over from one set of four to another set of four, we obtain a surface formed by the corners of the tetrahedra, each point of that surface corresponding to a point on the surface of the human body. We have then transformed the latter surface (in the empirical space) into a surface in the electrical image space. The line connecting two points on this surface is thus the lead vector for the two corresponding points on the human body. In general this vector evidently has an entirely different direction and magnitude from the line connecting the relevant contact points in the empirical space.

By way of illustration we shall take a simple two-dimensional example (*fig. 9*). It will be obvious that in this example both the imaginary heart vector and the two lead vectors each have only two components; we denote these respectively by X and Y, a_1 and b_1, and a_2 and b_2. The above-mentioned tetrahedron is reduced here to a triangle (whose position in the XY plane, with respect to the origin, is quite arbitrary). The figure shows how the magnitude and direction of the triangle sides can be derived from the values of a_1 to b_2, and also how one can find from the vectorcardiogram the conventional electrocardiograms that would have been obtained from the leads in question. It can be seen that the corresponding spikes in the conventional electrocardiograms do not occur simultaneously. Their sequence is determined from the direction in which the heart vector rotates.

A vectorcardiograph for clinical use

We shall now discuss a vectorcardiograph, built in Eindhoven for use in the Philips Health Centre, which is designed to operate with two lead systems and which can be moved from one place to another as required. The first lead system is one in which electrodes are placed on the left leg, the two arms

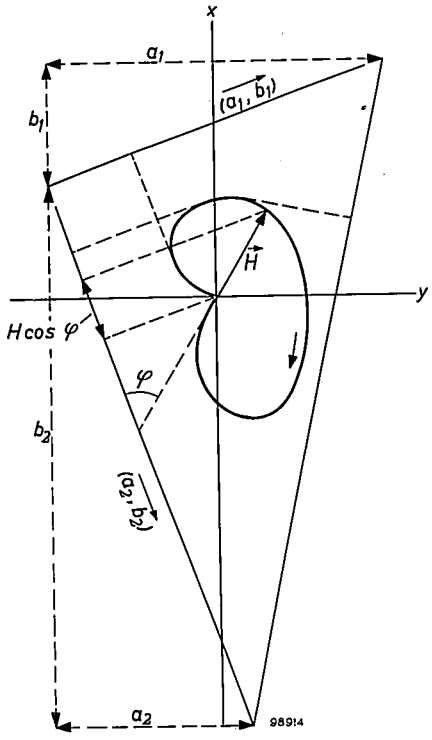


Fig. 9. Two-dimensional illustration of the relation between the heart vector $\mathbf{H}$ and the potential P_i produced between a given pair of electrodes. There are here two independent potentials P_1 and P_2. The appertaining leads (derivations) each give rise to a lead vector having components a_1, b_1 and a_2, b_2, respectively. These form a triangle (capable of parallel movement in the XY plane). The magnitude of P_1, for example, is found geometrically by projecting $\mathbf{H}$ to the relevant lead vector and by multiplying the length of the projection ($H \cos \varphi$) by the magnitude of this vector. It is seen that such a projection and hence P_i will reach its maximum value at different times for all three leads.

and on the middle of the chest in line with the top of the arm pits. The second system uses an additional electrode which is placed on the back at the height of the seventh vertebra, 2 cm left of centre. The apparatus (*fig. 10*) is equipped with a cathode-ray tube (screen diameter 13 cm), on which the spot can describe successively the frontal, the horizontal and the sagittal projection of the vectorcardiogram. A block diagram of the apparatus is shown in *fig. 11*. The potentials L, F, B and W which the left arm, the left leg, the chest and the back, respectively, have in relation to the right arm, are each separately amplified by the four amplifiers A_1 to A_4. The amplified signals are fed to the mixing

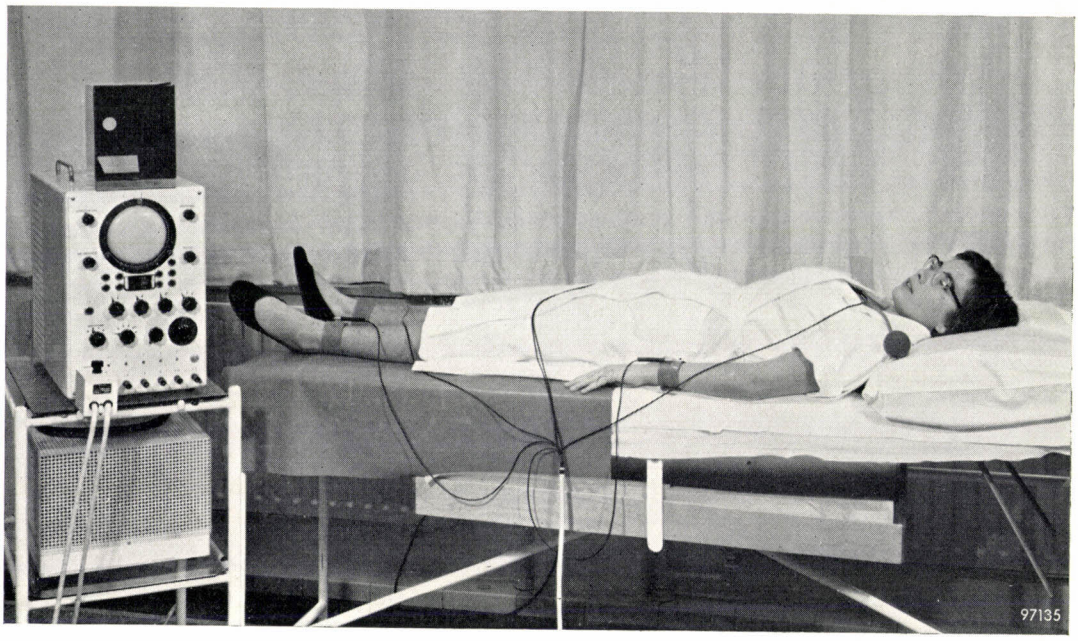


Fig. 10. The vectorcardiograph in use. Two cords can be seen connected to the instrument. One leads to the patient shown in the photograph, the other can be used for connecting a second patient whilst the recordings of the first are made. This makes the apparatus better suited for routine examinations. Each cable contains six conductors, five of which are connected respectively to the right leg (by which the patient is earthed), the right arm (the common point of all leads), the left arm, the left leg and the chest. The sixth can be connected to a dorsal electrode. In the photograph it is not in use (and for this reason is connected to the right arm). The dorsal and chest electrodes are applied by means of a sucker. The power-supply circuits are contained in a separate housing, shown below the oscilloscope in the photograph. The oscilloscope also houses the vectorcardiograph circuits proper. The bracket on which the camera can be mounted is folded back.

circuit M, which "computes" and adds the terms of the equations (2). Of the three signals produced by this circuit, which are proportional to the X, Y and Z components of the heart vector, two are applied, after further amplification by B_1 and B_2, to the horizontal and vertical deflection plates, respectively, of the oscilloscope, which thus always displays a projection of the vectorcardiogram. The beam modulator C contains three circuits. The first periodically suppresses the oscilloscope beam to produce time marks on the screen. The second varies the intensity of the beam with the writing speed; in the first place this prevents the beginning of the trace — where the spot remains for a relatively long period — from causing a large black mark on the film at that position, which would make it impossible to determine the origine of the vectorcardiogram; in the second place it prevents the slow parts of the trace from appearing exceptionally thick. The third circuit makes it possible to display only part of the projection while suppressing the remainder; this also enables the observer to determine the direction in which the vectorcardiogram is described.

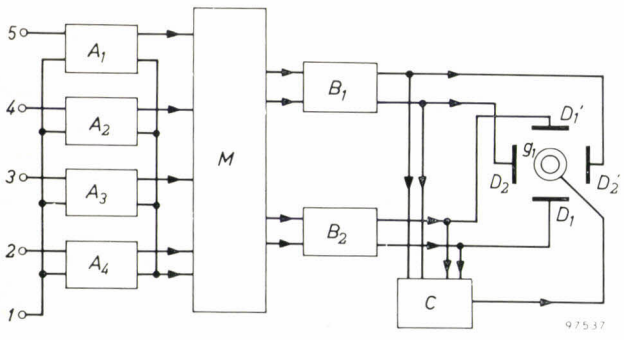


Fig. 11. Block diagram of vectorcardiograph. All potentials are measured with respect to the right arm (terminal *1*), namely those of the left arm, the left leg, the chest and possibly the back. A_1 to A_4 pre-amplifiers, M mixing circuit, B_1 and B_2 output amplifiers, D_1-D_1' and D_2-D_2' deflection plates of oscilloscope, g_1 CRT control grid, C beam modulator.

We shall now discuss each of these components in turn, after which we shall touch on ways and means of suppressing external interference.

Pre-amplifiers

The pre-amplifiers in a vectorcardiograph are required to amplify the potential difference of two points, neither of which are at earth potential, in such a way that the average of the potential on these points in no way affects the output signal. This requirement, which applies equally to a conventional electrocardiograph, is necessary not only because the action of the heart causes the potential of the

contact points to deviate from earth potential, but also because of the occurrence of interfering signals (hum) which cause the potential of the whole body to differ from zero. Further, a high input impedance is required because 1) the interior of the body cannot be regarded as a voltage source of low internal resistance, and 2) a current flowing from the body to an electrode plate encounters a resistance of several kΩ in passing the skin, even when the skin is treated with an electrode paste [6]). Finally, the amplifier must not amplify constant voltages, arising among other things from the use of electrode paste, although the lower limit of the frequency characteristic should not be higher than approximately 0.1 c/s.

The circuit designed for the amplifiers A_1 to A_4 (fig. 11) and which meets the requirements mentioned to a high degree, is shown in *fig. 12b*. The first stage (under the dashed line) can be regarded as derived from a simple pentode difference-amplifier (fig. 12*a*) in the following way: 1) the pentodes are each replaced by a series (cascode) arrangement of two triodes; 2) the common cathode resistor is replaced by a triode *III*; 3) the circuit supplying the grid bias for the triodes *Ib* and *IIb* (and which is comparable with the battery supplying the screen-grid voltage in the pentode amplifier) consists of a triode *IV* connected as a cathode follower. This stage has a rejection factor of more than 10^4, that is to say, the change in the *average* potential of g_1 and g_1' must be at least 10^4 times larger than the potential *difference* of g_1 and g_1' in order to make the same contribution to the output signal. The average potential of $\frac{1}{2}(V_{g1} + V_{g1}')$ can thus vary within a wide range compared with that in which $V_{g1} - V_{g1}'$ varies, without this being noticeable in the output signal [7]).

The second stage (above the dashed line in fig. 12*b*) is a simple triode difference-amplifier. The gain of this stage can be controlled with the variable resistor R. The two stages are coupled by capacitors of 2 μF, and 2.2 MΩ grid resistors are used for the tubes in the second stage.

[6]) The impedance present between the interior of the body (which is a relatively good conductor) and an electrode plate can be represented by a resistor R and a capacitor C connected in parallel. Measurements in which the plate was applied to the body with the aid of electrode paste (this always contains an electrolyte and sometimes too an agent for softening the horny layer of the skin) and in which the skin was properly softened, showed for a plate of 9 cm²: R = 2.5 kΩ and C = 0.3 μF, and for a plate of 36 cm²: 1.5 kΩ and 0.75 μF. Less careful application of the plate made R in some cases ten times higher. The resistance of the interior of the body was found to be 260 to 270 Ω from wrist to wrist.

[7]) See G. Klein, Rejection factor of difference amplifiers, Philips Res. Repts. **10**, 241-259, 1955.

The total gain of the amplifiers A_1 to A_4 is approximately 5000×. The total rejection factor is approximately equal to that of the first stage and hence very high. The circuit of this amplifier stage offers

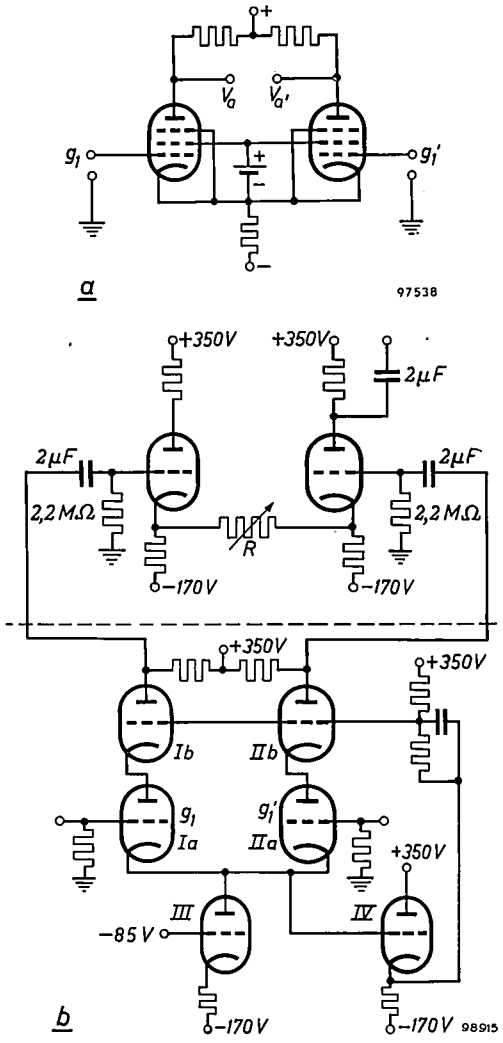


Fig. 12. *a*) Diagram of a simple pentode difference-amplifier. If both pentodes and their anode resistors are completely identical, the output signal (V_a–V_a') is proportional to (V_{g1}–V_{g1}') and independent of $\frac{1}{2}(V_{g1} + V_{g1}')$.
b) Circuit diagram of the pre-amplifiers (A_1-A_4 in fig. 11). The first stage (below the dashed line) can be regarded as derived from the pentode difference-amplifier (*a*), the pentodes being replaced by cascodes (i.e. the triode-pairs *Ia* and *Ib*, and *IIa* and *IIb*, connected in series), the cathode resistor being replaced by the triode *III*, and the battery supplying the screen-grid voltage by the circuit of triode *IV*. The latter circuit ensures that the grid potentials of triodes *Ib* and *IIb* closely follow the changes in the cathode potentials. This stage has a high rejection factor, although no selected tubes or resistors are employed. The second stage (above the dashed line) is a simple triode difference-amplifier, the gain of which can be controlled by the variable resistor R.

the very important practical advantage that the above-mentioned rejection factor is obtained without the use of specially selected tubes or resistors. Moreover, because of this very high rejection factor, no particularly high demands need be made with respect to the constancy of the supply voltage.

We shall give here some further details of the considerations that led to the circuit design of the first stage.

Calculations show that a difference amplifier of the type in fig. 12*a* can be given a large rejection factor by using tubes having a high amplification factor μ and by choosing the magnitude of the common cathode resistance such that the product of this resistance and the average transconductance (slope) of the tubes is of the same order of magnitude as μ. Now the effective μ value for a cascode arrangement is equal to the product of the μ values of the individual tubes, provided the grid potentials of the "upper" tubes differ by a constant amount from the cathode potentials of the "lower" tubes (and hence follow the potential changes of the latter). With such a cascode arrangement, then, a very high effective μ can be achieved. The common cathode resistance can be given a correspondingly high effective value by using a triode for this purpose, or perhaps another cascode set-up. In this way an effective value of 10 to 100 MΩ can easily be obtained. This subject is dealt with at length in the article quoted under [7]).

Mixing circuit and output amplifiers

The linear equations that express (proportionately) the components of the heart vector in terms of the potentials appearing between the electrode pairs were chosen as follows for the system with three leads (four electrodes):

$$X = \quad 8B + 54L + 16F,$$
$$Y = -\ 6B - 10L + 26F,$$
$$Z = -40B + 12L - 26F,$$

and for the system with four leads:

$$X = \quad 3B - 9W + 59L + 15F,$$
$$Y = -\ 2B + 8W - 11L + 26F,$$
$$Z = -13B + 51W - 2L - 27F,$$

where L, F, B and W represent the earlier-mentioned potential differences. The correct fractions of the signals from the pre-amplifiers are obtained, as we have seen, by means of potentiometer circuits. With the aid of a number of switches (on a single spindle) it is possible to switch between the coefficients of one system of equations and those of the other.

All signals having a positive coefficient in a particular equation are thereupon fed to an adding circuit; those having a negative coefficient likewise. These adding circuits are equipped with triode cathode followers (three at most) and with a common anode resistance (*fig. 13a*). Since a single projection calls for no more than two summations at a time, only four such adding circuits are required. (One circuit in each pair contains a triode whose control grid is maintained at earth potential and whose cathode resistance is variable (fig. 13*b*); this makes it possible to shift the patterns on the oscilloscope screen.) Of each pair of adding circuits, then, one supplies the absolute value of the sum of the positive terms, and the other that of the sum of the negative terms of the equation concerned. A signal proportional to the difference of both sums

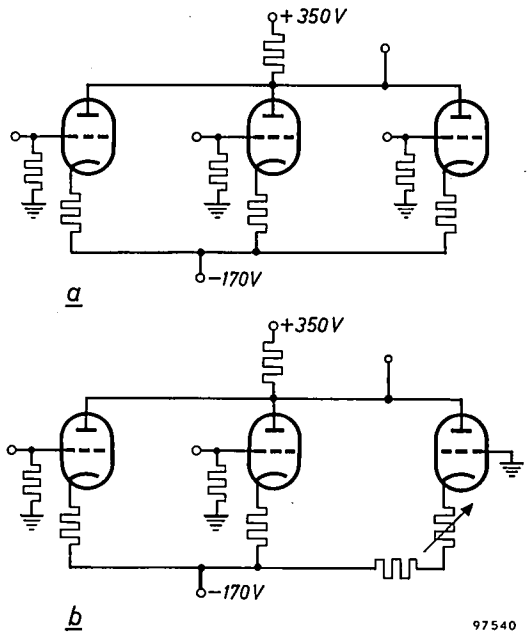


Fig. 13. *a*) Adding circuit with three triodes connected in parallel, with a common anode resistor and identical cathode resistors. The signals to be added are applied to the three grids; the sum signal is proportional to the variations of the anode voltage. *b*) Similar circuit in which one triode has an earthed grid and a variable cathode resistor. By means of this triode a constant voltage can be applied to the sum signal, making it possible to change the position of the pattern on the oscilloscope screen without affecting the shape of the oscillogram.

is obtained by applying the summed signals to one of the output amplifiers. These are again designed as difference amplifiers (*fig. 14*). The amplification of the difference signal (max. 60×) can be reduced continuously by approximately a factor of 10. A large rejection factor is not necessary here.

The frequency response of the whole apparatus is determined at the low-frequency side by that of the pre-amplifiers together with the mixing circuit, and at the high-frequency side by that of the output amplifiers. The lower limit of the frequency band passed by the apparatus is mainly determined by the value of the capacitor that couples the pre-amplifier to the potentiometers in the mixing circuit, and by the resistance of these potentiometers (2 μF and 1.2 MΩ, respectively). The coupling between the two pre-amplifier stages has an even higher RC value (2 μF × 2.2 MΩ) and is of secondary significance. The frequency at which the gain has dropped to 70% ($1/\sqrt{2}$) is approximately 1/12 c/s. The upper limit of the frequency band is approximately 100 c/s. The choice of this upper limit is determined partly by the consideration that a

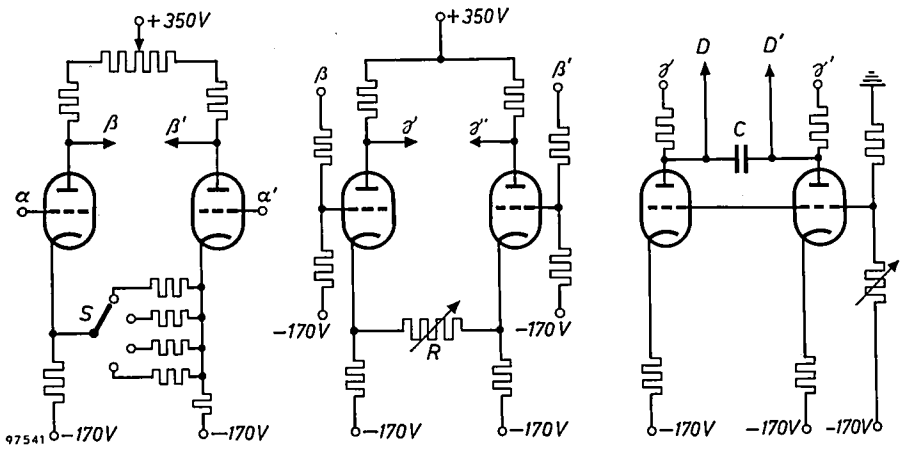


Fig. 14. Circuit diagram of output amplifiers. The first stage (left) is connected to the second stage (centre) at the positions β and β', the second stage to the output stage (right) at the positions γ and γ'. Both stages are difference amplifiers. The input terminals of the first stage are the points α and α'. The gain can be varied in four steps by switch S (approximately as 1:2:4:8) and continuously with the variable resistor R. The function of the output circuit is to enable the output signal of the second stage, appearing at (γ, γ') where the average potential is about +230 V, to be applied without any attenuation to the deflection plates D, D', where the average potential is that of earth. The capacitor C together with the resistors of the output circuit determines the upper limit of the frequency band passed by the amplifiers.

higher limiting frequency is unnecessary for cardiography, and partly by the fact that a higher value would increase the influence of interfering potentials from other muscles.

Beam modulator

The beam modulator consists, as mentioned, of three parts: one which periodically suppresses the beam, one for varying the intensity with the writing speed, and one which enables the projection to be only partly displayed. The latter two parts have one section in common. A diagram of the whole modulating system is given in *fig. 15*. The common section referred to is found in (A) above the upper dashed line. It consists of two double triodes connected in parallel and each with a common cathode. The grids of each double triode are connected via coupling capacitors of 1500 pF to the ends of the output amplifiers, that is to say each grid is connected to one of the four deflection plates of the oscilloscope. Since the changes in the potential on the plates of one pair are always equal but opposite, this also applies to the signals fed to the grids of the relevant pair of triodes. The latter signals are approximately proportional to the time derivatives of the potentials on the deflection plates, i.e. they are approximately proportional to the horizontal and vertical components, respectively, of the writing velocity of the spot on the oscilloscope screen. The negative signal from each pair rapidly cuts off the relevant triode, whilst the positive signal determines the current through the common cathode resistor. The output signal from this part of the circuit is therefore approximately proportional to the sum of the absolute values of the horizontal and vertical velocity components of the spot on the screen, and thus varies in approximately the same way as this velocity. Through the circuit in B this signal, which is always positive, is fed to the output (g_1, in D), i.e. to the control grid of the cathode-ray tube. The intensity of the electron beam is consequently higher when the writing velocity is greater, so that the line traced has roughly uniform thickness.

Also connected directly to the output is the circuit shown in C. This is simply a multivibrator, whose function is to interrupt the electron beam periodically and thus to provide the time marking. If the switch S_1 is in position N, the multivibrator is out of operation, and an uninterrupted trace is obtained. In the other case the beam is interrupted 200 times per second.

The third function of the beam modulator, which is to suppress the beam during a certain part of the cardiac cycle, is performed by the circuit shown in E and F. F contains two monostable multivibrators [8]) with variable pulse width. The pulse width of the one on the right governs the length of the time interval during which the electron beam is passed or

[8]) A monostable multivibrator has one stable and one quasi-stable state. A trigger pulse of the appropriate sign drives the circuit into a state in which the tube that was conductive is cut off and the other made conductive; after a predetermined period the circuit returns to the original state. The output signal is a rectangular voltage pulse whose duration is equal to the above predetermined period.

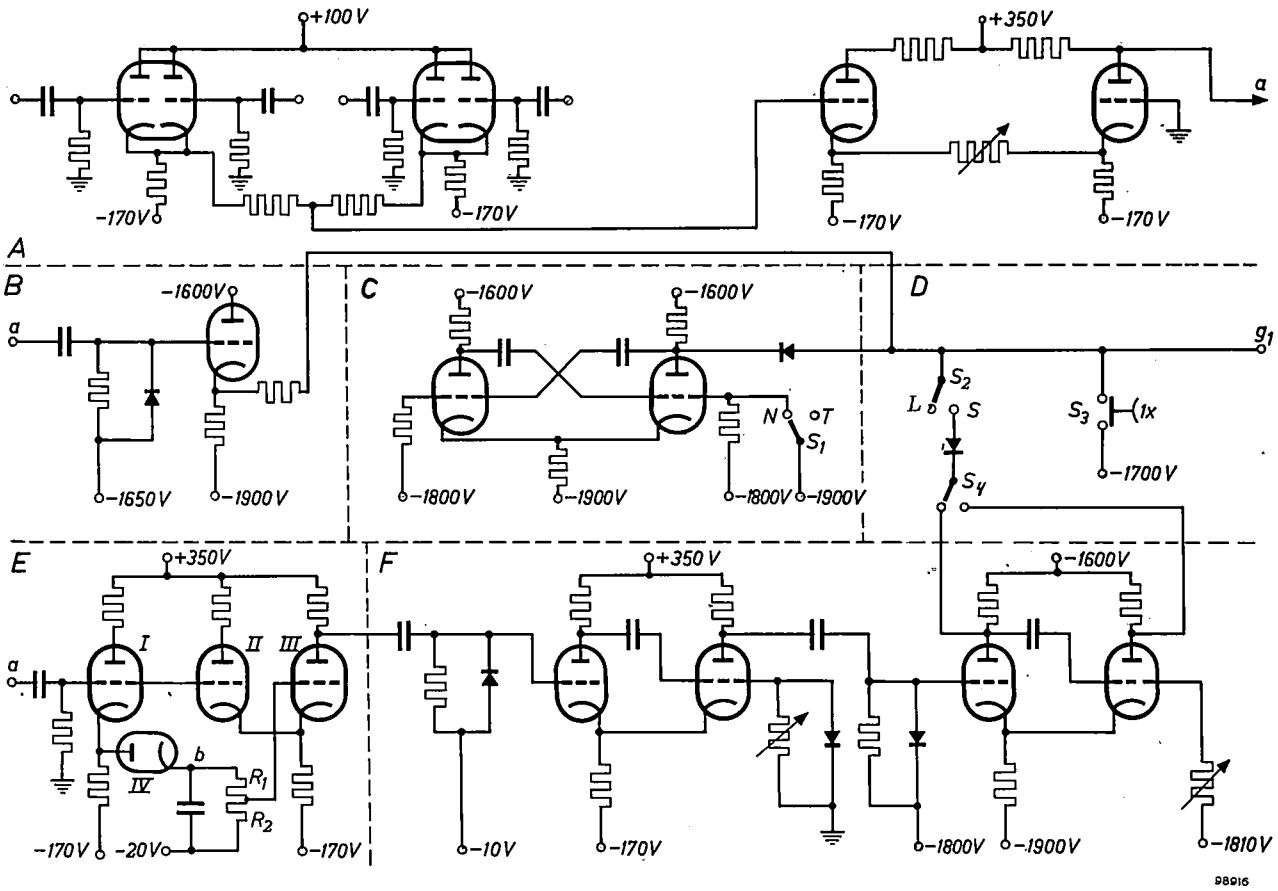


Fig. 15. Circuit diagram of beam modulator which *a*) makes the thickness of the beam vary with the writing speed, *b*) enables a certain part of the figure to be suppressed, and *c*) periodically interrupts the beam. The circuits fulfilling the first two functions have a common section (*A*), the signal from which is approximately proportional to the sum of the absolute values of the horizontal and vertical components of the writing speed (i.e. proportional to the sum of the derivatives — obtained via the coupling capacitors — of the potentials of the positive plate of each pair of deflection plates). This signal is amplified by the two triodes shown on the right of section *A*, and serves as the input signal for section *B* and also for section *E*. The first simply applies the signal to the output of the beam modulator (g_1, section *D*), i.e. to the control grid of the cathode-ray tube. The purpose of the diode (section *B*) is to prevent the triode grid potential from falling below the quiescent value. This circuit performs the function mentioned under *a*). The circuit in section *E* delivers a trigger pulse once per heart beat to the first monostable multivibrator [8]) in section *F*, exactly at the moment at which the incoming signal has maximum amplitude. The operating period of both monostable multivibrators (*F*) is variable. The period (pulse width) of the second determines the length of time the beam shall be suppressed (or passed, depending on the position of S_4), whilst that of the first monostable multivibrator determines the moment at which this time interval begins. Switch S_2 makes and breaks the connection with the output. The circuit in section *C*, a multivibrator, supplies the square-wave voltage for periodically interrupting the electron beam (function *c*). The multivibrator oscillates when S_1 is in position *T*. The diodes in sections *C* and *D* break the connection with g_1 as soon as the potential at the output of the circuit concerned rises above that produced by section *B* on g_1. S_3 is a push-button for manually suppressing the beam.

suppressed (depending on the position of switch S_4). The pulse width of the other governs the moment at which this time interval begins. During each cardiac cycle this monostable multivibrator is triggered at corresponding moments by the circuit shown in *E*. The signal applied to this circuit is the output signal from the circuit in branch *A*, and is thus proportional to the sum of the absolute values of the two velocity components of the light spot. For each heart beat this signal consists of a number of peaks, which are always positive and one of which is usually considerably larger than the others. This large peak, the top of which coincides with the moment at which the writing velocity in the *QRS* loop is at maximum, makes the diode *IV* conductive, thus increasing the potential on point *b*, and hence also (to an extent determined by potentiometer R_1-R_2) the potential on the grid of triode *III*. As a result triode *III* passes a fairly high current and triode *II* is cut off. Owing to the very high *RC* constant of the circuit in question (3.5 sec) this situation is maintained long enough to prevent the diode *IV* or the triode *II* from being made conductive by any of the smaller peaks following upon the maximum peak during the same heart beat, or by those preceding this peak during the following heart beat. Only the large peak of the successive heart beat can make these tubes conductive, and thus only at the moment of its arrival does the circuit deliver a trigger pulse to the first monostable multivibrator in *F*.

Other circuits

Finally, some remarks about the power-supply circuits, the cathode-ray tube and the calibrating oscillator. As regards the first of the two supply circuits it need only be mentioned that its output voltage is 520 V, the potential of one terminal being —170 V and that of the other +350 V. The output voltage (520 V) and the potential on the negative terminal (—170 V) are stabilized. A 10% mains fluctuation changes the output voltage by approx. 50 mV and the potential on the negative terminal only by about 3 mV. This circuit supplies not only the anode voltage for the various components of the vectorcardiograph, but also the filament current for the pre-amplifier tubes and for various tubes in the supply circuits. For this purpose the filaments of these tubes (all type UCC 85; heater current 100 mA) are connected in series.

The other supply circuit, which among other things provides the high tension for the cathode-ray tube, is shown in *fig. 16*. Here again, use is made of the properties of series-connected triodes. This circuit, details of which are given in the subscript to the figure, can only be employed, however, in applications where the current drain is low and fairly constant.

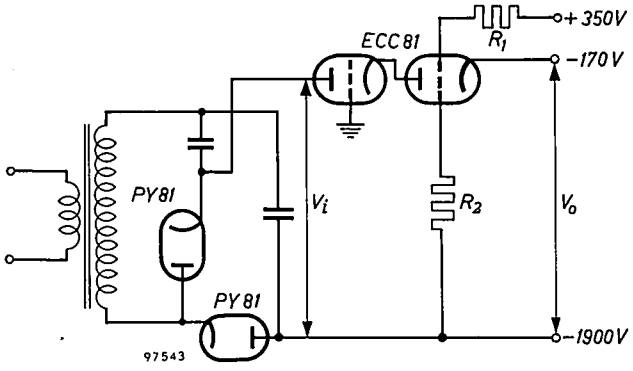


Fig. 16. Simplified diagram of the circuit that supplies the stabilized direct voltage for the cathode-ray tube and for certain components of the beam modulator. The output voltage V_o is very well stabilized with respect to variations in V_i (i.e. the mains voltage), but less so with respect to load variations. The proportionality factor between the variations in V_i and V_o for this circuit is $\mu R_1/(R_1+R_2)$, where μ is the effective amplification factor of the cascode. Since $\mu = 3600$ and $R_1/(R_1+R_2) = \frac{1}{4}$, this factor is 900. The stabilized voltage for the cascode grids is supplied by the other power-supply circuit of the vectorcardiograph (+350 V, —170 V). The voltage V_i is obtained from a high-tension transformer (secondary voltage approximately 800 $V_{r.m.s.}$) and a voltage-doubling rectifier circuit.

The cathode-ray tube is of the type DP 13-2 with both blue and yellowish-green fluorescence. The blue component has a short afterglow, which is an advantage when photographing the screen, and the yellowish-green component has a long afterglow, which is an advantage for visual observation.

The calibration-voltage generator delivers a sinusoidal signal of variable amplitude (max. 1 $V_{r.m.s.}$ approx.) and 25 c/s frequency. With the voltage output from this oscillator it is possible to check not only the operation of the amplifiers (for both gain and rejection factor) but also all connecting leads, including the lead used for earthing the patient (see below).

Suppressing of external interference

The chief form of external interference experienced in cardiography is induction from the electric mains (hum). In the first place this causes the potential of the body as a whole to fluctuate, and secondly it gives rise to potential differences between the various points on the surface of the body. The average potential difference between the body and earth can be reduced to a few mV at the most by earthing the patient. (We have in mind a patient in a normally equipped laboratory or hospital ward, but not of course a case where the patient is very close to cables carrying the mains voltage.) The limiting factor here is the resistance of the skin which, as mentioned earlier, can amount to several kΩ. In the apparatus described the patient is earthed via the right leg. Owing to the high rejection factor of the pre-amplifiers, the hum signal after earthing is not troublesome. The potential difference induced between the various parts of the body is found to be 20 to 30 μV between the wrists of the left and right arm when the patient is earthed via the right leg. This is just below the nuisance limit, and therefore it is possible with this apparatus to record vectorcardiograms at any required location without special precautions being necessary.

The influence of a second kind of external interference, namely microphony caused by mechanical vibrations, is eliminated by spring-suspension of the pre-amplifiers.

Practical significance of vectorcardiography

Although vectorcardiography is still to some extent in the stage of research and is not yet in general clinical use, it has already proved its great practical value as a method of examining the heart.

A very useful property of the vectorcardiogram is that its shape can fairly easily be correlated with the anatomy of the human heart. This means that it is comparatively simple to learn to interpret a vectorcardiogram, whereas for interpreting the results of a conventional electrocardiographic examination it is necessary to memorize a large number of patterns whose relation to the anatomy of the heart is by no

means obvious. Experience has shown that a physician who has first practised vectorcardiography and then proceeds to study ordinary electrocardiography is able to grasp this relationship more readily and thus to interpret conventional electrocardiograms accurately sooner than others; there is thus no doubt that vectorcardiography also possesses didactic value. In some pathological cases the electrical phenomenon is so capricious as to defy interpretation by the conventional electrocardiographic method. In such cases the vectorcardiogram may often provide the answer. Certain details appearing in the vectorcardiogram are often not sufficiently evident in a conventional electrocardiogram to be detected.

In connection with simplicity of interpretation, moreover, it may be added that it is easier to interpret shapes than to read graphs.

An important particular that can be directly deduced from a vectorcardiogram is the sense of rotation of the heart vector. True, this can also be derived from ordinary electrocardiograms, but only by ascertaining the moment at which corresponding peaks occur in the various leads (cf. fig. 9).

Because of the ease of interpretation it offers and the reasonably high percentage of cases in which the heart vector is recorded with good accuracy, vectorcardiography in its present form can already be used as a routine screening test and is in fact so used in Eindhoven. It offers the physician the possibility of rapidly determining the location and the nature of a disorder.

Since vectorcardiography looks at the heart, as it were, from a distance and provides no information on an individual small part of the heart, it is not to be expected that it will ever entirely supersede conventional electrocardiography, which can provide detail information. The one method will supplement the other, in the sense that conventional electrocardiography will continue to provide additional detailed information.

Summary. The electrical action of a contracting muscle fibre is equivalent to that of a dipole and can be represented by a vector. The same applies to an entire muscle, such as the heart, insofar as the electrical phenomena are observed at a large distance from the source. During the heart beat the heart vector changes in magnitude and direction. Since the human body is a conductor, the electrical action of the heart manifests itself in the form of potential differences between various points on the surface of the body; these potential differences show marked variations as a function of time, and recorded they produce the conventional electrocardiogram. By applying electrodes to at least four positions on the body and combining the independent potential differences in the correct manner — the relation between the components of the heart vector and each of these potential differences is a linear one — the 3-dimensional curve can be found that represents the variation of the heart vector with time (called the vector-electrocardiogram). In the vector-electrocardiograph described, two of the three mutually perpendicular components of the heart vector are "computed" from the potential differences mentioned and proportional voltages are applied to the horizontal and vertical deflection plates of a cathode-ray oscilloscope. Successive recordings are thus made of the frontal, horizontal and sagittal projections of the vectorcardiogram. Time marks are obtained by periodically interrupting the electron beam; the sense of rotation of the curve is found by breaking off the recording at a certain moment. The use of difference amplifiers having a high rejection factor makes the apparatus very insensitive to external interference that causes equal fluctuations of the electrode potentials (hum). Since the form of vectorcardiograms can be fairly simply correlated with the anatomy of the heart, it is easy to learn to interpret them. Even in its present form vectorcardiography can already be put to good use for didactic purposes and as a routine screening test. Moreover the significance of some details in a conventional electrocardiogram sometimes only becomes clear after comparison with a vectorcardiogram. The nature of the method justifies the conclusion that vectorcardiography will prove a useful supplement to conventional electrocardiography, but will not supersede it.

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 Laboratories, Eindhoven, Netherlands.

2638: H. G. van Bueren: Versterkers voor licht en radiostraling uit de wereldruimte (Ned. T. Natuurk. **24**, 189-205 and 213-228, 1958, Nos. 8 and 9). (Amplifiers for light and radio waves from space; in Dutch.)

Survey of a number of modern detectors of light and microwaves for use in experimental astronomy. While "classical" light detectors such as the photographic plate and the photocell have adequate *absolute* sensitivity, it is nevertheless worth while to investigate any methods which might shorten observation times or improve resolution, particularly for large faint objects or objects whose brightness varies rapidly. Radio waves from outer space are very weak and have the character of noise; their detection and investigation requires receivers of very low noise level. Modern developments in microwave techniques and solid state physics have led to totally new approaches to the amplification of radiation. Among the new devices discussed here are masers, parametric amplifiers and various types of light amplifiers.

2639: J. Davidse: Versuche über die Anpassung des NTSC-Farbfernsehsystems an die europäische 625-Zeilen-Norm (Nachr.tech. Z. **11**, 461-466, 1958, No. 9). (Investigation of the application of the N.T.S.C. colour-television system to the European 625-line standard; in German.)

Investigations into the factors which determine system parameters of transmission systems for colour television according to the American N.T.S.C. system. The work is particularly directed to the application of the N.T.S.C. system to the 625-line television standard (Gerber-standard) used in certain countries on the continent of Europe. The choice of the two colour signals and their bandwidths and the choice of the sub-carrier frequency are discussed. Experiments on loss of sharpness as a result of bandwidth restriction of the colour information and on crosstalk phenomena show that the optimum choice of the system parameters is primarily governed by the statistical properties of the colour signal.

R 364: K. F. Niessen: Distribution of magnetic domains between the two phases in a single-crystal flat disk of iron (Philips Res. Repts. **14**, 101-110, 1959, No. 2).

A single-crystal oblate spheroid of iron is considered whose plane is a (001) plane of the crystal. It is subjected to a magnetic field H lying in the first quadrant of this plane, making an angle with the second of the two preferential axes [100], [010] which lie also in the plane. The total number of magnetic domains is called N, all being of the same moment m, whereas Nm is supposed to be known. The ratio in which these domains are distributed between the preferential directions of the crystal depends on the strength H of the external field. The graphical determination of this ratio is the aim of the present paper. It is compared with the analytical determination given by Néel and by Lawton and Stewart.

R 365: K. Böke, J. B. M. Spaapen and N. B. Speyer: Diffusion capacitance in transistors (Philips Res. Repts. **14**, 111-122, 1959, No. 2).

Capacitance measurements have been performed with transistors at different temperatures, voltages and frequencies. Calculations of the diffusion capacitance (= injection or storage capacitance) have been carried out, taking into account the influence of the second junction. Theory and experiment both indicate that this influence can be very large. The calculations are in qualitative and quantitative agreement with the measurements.

R 366: W. Ch. van Geel and C. A. Pistorius: Current-time relationship in the forward direction of electrolytic rectifiers (Philips Res. Repts. **14**, 123-131, 1959, No. 2).

The shape of the forward current as a function of time is given for an electrolytic rectifier, if an alternating rectangular voltage is applied. The loops in the current-voltage characteristic, which occur when a sinusoidal voltage is applied, are explained. It appears that the forward current decreases strongly with increasing blocking voltage.

The forward current through the oxide layer hardly depends on the thickness of this layer. A qualitative explanation of the observed phenomena is given.

R 367: G. Brouwer: Three-dimensional electric-circuit model of the high-frequency phenomena in a junction transistor (Philips Res. Repts. **14**, 132-142, 1959, No. 2).

In order to give a complete explanation of the high-frequency behaviour of a junction transistor both minority- and majority-carrier flow must be taken into account. Owing to the large number of complicated boundary conditions, it is not possible to give the solution in a closed mathematical form. The linearized problem, corresponding to small-signal operation of the transistor, may be solved with the aid of an electric-circuit model. The high-frequency current gain of the transistor may be derived from model experiments.

R 368: J. J. Scheer and P. Zalm: Crystal structure of sodium-potassium antimonide (Na_2KSb) (Philips Res. Repts. **14**, 143-150, 1959, No. 2).

X-ray analysis of Na_2KSb — a photoemissive material discovered by Sommer — has led to the determination of its crystal structure. The unit cell is described by the space group $Fm3m\text{-}O_h^5$ with four antimony atoms at $(0,0,0;\ 0,\frac{1}{2},\frac{1}{2};\ \frac{1}{2},0,\frac{1}{2};\ \frac{1}{2},\frac{1}{2},0) + 0,0,0$; four potassium atoms at $(\quad) + \frac{1}{2},\frac{1}{2},\frac{1}{2}$ and eight sodium atoms at $(\quad) + \frac{1}{4},\frac{1}{4},\frac{1}{4}$ and $\frac{3}{4},\frac{3}{4},\frac{3}{4}$. The crystal structure of Na_2KSb has a great resemblance to that of Cs_3Sb and differs strongly from that of Na_3Sb and K_3Sb which both crystallize in the Na_3As type.

R 369: N. C. de Troye: Classification and minimization of switching functions (Philips Res. Repts. **14**, 151-193, 1959, No. 2).

Electronic computers consist of a large number of switching elements, of which there are relatively few types, which together are capable of handling or forming a large number of conditions. A special switching technique, with the aid of diodes, is now being used for those parts of the machine where extremely short switching times are required. For several reasons it is advisable to keep the number of diodes as small as possible. To study the various conditions to be realized with diode circuits, Boolean algebra may be profitably applied. It may be demonstrated that two forms of notation in Boolean algebra, viz. the minimal sum of products and the minimal product of sums, are of particular importance in diode-circuit configurations. Several authors have attempted to arrive at these two forms of notation with varying degrees of success. This paper (thesis, Amsterdam 1958), too, is an attempt to find from a given Boolean function — also called switching function in view of its application in computer techniques — either the minimal sum of products or the minimal product of sums. It is demonstrated that it is possible to transform any given switching function into a matrix containing only the elements 0 or 1. The number of elements 1 in the various submatrices indicates whether a simplified notation of the switching function is possible. The possibility of easily finding the prime implicants of the switching function is likewise shown. These prime implicants can then be used to determine the minimum sum of products. It is found that this process can be carried out by means of electronic computers. The number of switching functions of n variables is 2^{2^n}. As is demonstrated, it is not necessary to determine the minimum sum of products for all these switching functions if the concept of equivalence class (i.e. the set of all switching functions that are invariant as regards permutation and negation of variables) is introduced. Every equivalence class has a representative and it is only of this representative that the minimal sum of products has to be obtained. Determining the equivalence class for any given switching function of 3 or 4 variables is a relatively simple matter.

R 370: W. L. Wanmaker, M. L. Verheijke and W. Parchen: Influence of the "dope" on the reduction rate of tungsten trioxide by hydrogen (Philips Res. Repts. **14**, 201-206, 1959, No. 3).

The influence of water vapour, layer thickness and presence of additives ("dopes"), e.g. K silicate and $K_4SiW_{12}O_{40}$, on the reduction rate of WO_3 is studied. K silicate and $K_4SiW_{12}O_{40}$ increase the reduction rate, but only if thick layers of WO_3 are reduced. The presence of water vapour imparts a retarding action to the reduction, which is less pronounced with the doped products.

R 371: H. U. Harten: Influence of the ambient atmosphere on the surface recombination of silicon (Philips Res. Repts. **14**, 207-210, 1959, No. 3).

From measurements of the surface photo-voltage it follows that the surface potential of silicon can be altered, similarly to that of germanium, over a wide range by chemical surface treatments and over a smaller range by the ambient atmosphere. Simultaneous investigation of the surface recombination

shows this process to be determined chiefly by recombination centres of the "Hall-Shockley-Read" type.

R 372: M. Avinor: Edge photoconductivity of cadmium sulphide (Philips Res. Repts. **14**, 211-214, 1959, No. 3).

It is shown that the characteristic photoconductivity peak of single crystals of cadmium sulphide at 515 mμ is not due to band-band transition. An additional photoconductivity peak is observed at 500 mμ by activation with silver in excess of co-activator. Superlinearity at room temperature was also observed.

R 373: Nathan Spielberg: Intensities of radiation from X-ray tubes and the excitation of fluorescence X-rays (Philips Res. Repts. **14**, 215-236, 1959, No. 3).

The intensity of fluorescence X-rays is discussed in terms of the efficiency of conversion of given primary wavelengths to fluorescence radiation and the spectral distribution of the primary radiation from the X-ray tube. For single-element specimens and major constituents of multiple-element specimens in the geometries normally used in modern spectrochemical analysis, the conversion efficiency varies slowly with primary wavelength if the exciting wavelength is not too far from the specimen absorption edge. In these cases the fluorescence intensity is affected primarily by the radiation flux, expressed in photons/second, from the exciting X-ray tube, and not so much by its wavelength distribution. For minor constituents of the specimen and/or when the exciting radiation is far from the specimen absorption edge, the conversion efficiency varies with the cube of the primary wavelength. The expected spectral intensity distribution, both in the characteristic and the continuous spectrum, from an X-ray tube target is discussed as a function of the target material. For the X-ray spectrochemical analysis of light elements (lighter than sulphur), the *L*-series radiation from a Mo or Ag target tube may lead to a large increase in fluorescence intensity. Experimental results confirming certain aspects of the theory are presented, and the need for further experiments, particularly with respect to soft X-rays, is pointed out.

R 374: J. Hornstra and P. Penning: Birefringence due to residual stress in silicon (Philips Res. Repts. **14**, 237-249, 1959, No. 3).

Rapid cooling of large silicon crystals gives rise to plastic deformation, as can be deduced from the increase of the etch-pit density. Additional information is obtained from a study of the residual stress by observation of the birefringence with infrared light. When slices are cut from the quenched cylindrical crystal the stresses change. The calculated stresses are corrected for this effect. Tensile stress is measured by compensation with external pressure. Also the optical phase difference is determined and, by comparison of the two, the stress-optical constant. The authors found a smaller value than Giardini; the difference, however, need not be significant. The experiments strongly suggest that a very large part of the thermal stress is released by plastic deformation, when the initial temperature is above 1300 °C. Residual stresses are also found in bars that have been bent at 750 °C. From the stress pattern the stress distribution during plastic flow can be derived. This gives directly the relation between flow rate and stress. In some cases the relation $\dot{\varepsilon} \propto \sinh(\sigma/\sigma_0)$, introduced by Van Bueren, was found; but often the stress pattern was more complicated due to the beginning of work hardening.

R 375: N. C. de Troye: Classification and minimization of switching functions (Philips Res. Repts. **14**, 250-292, 1959, No. 3).

Continuation of **R 369**.

VOLUME 21, 1959/60, No. 2 pp. 41-72 Published 16th December 1959

Philips Technical Review

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

A WIDE-BAND TRIODE AMPLIFIER WITH AN OUTPUT OF 10 W AT 4000 Mc/s

by J. P. M. GIELES and G. ANDRIEUX *). 621.375.2.029.6:621.385.3.029.6

This article describes an amplifier developed for use with the microwave disc-seal triode EC 59, dealt with earlier in this journal. The article follows upon an earlier one concerning an amplifier for the EC 157 triode, designed for lower outputs.

The disc-seal triode type EC 59 was developed for use in microwave radio links in which the power delivered by the EC 157 is inadequate [1]). In view of its higher output, and also because of some constructional differences between the two tubes, the EC 59 cannot be used in the amplifier designed for the EC 157 [2]). This article describes the construction and properties of an amplifier specially developed for use with the EC 59. Since both amplifiers obviously resemble each other in many respects, we shall make frequent reference to the earlier article [2]), denoted here as I.

Principal differences between the triodes EC 59 and EC 157

Fig. 1 shows the above two tubes side by side. The chief difference between them is that the EC 59 can deliver a power of more than 10 W, whilst the output of the EC 157 is only 1.5 W. For this reason the anode of the EC 59 is *water-cooled*, and this called for a radical modification in the design of the amplifier. Another important difference between the two tubes is the size of the cathode holder. In the EC 59 its length has been reduced and its diameter increased. This entailed a different construction in order to match the tube properly to the waveguide.

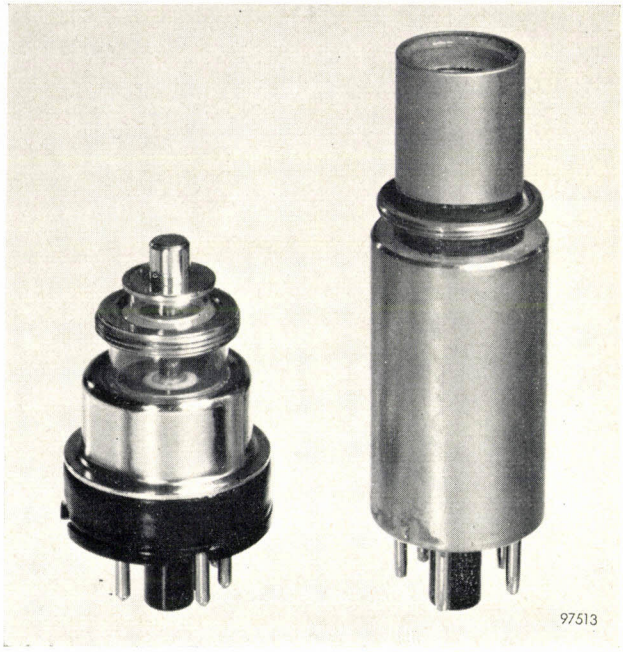


Fig. 1. Disc-seal triodes EC 157 (air-cooled) and EC 59 (water-cooled), both for 4000 Mc/s.

The input circuit

Fig. 2 shows an equivalent circuit for the input impedance of a disc-seal triode (see also fig. 3 in I). R_{gk} is the input resistance measured directly between grid and cathode, C_{gk} the capacitance between these electrodes and L_k the inductance of the cathode base, the cathode bush and the cathode disc. Parallel with this is the capacitance C_d between grid disc and cathode disc. In the two tubes EC 157 and EC 59, R_{gk} is about 60 Ω and C_{gk} about 1.6 pF.

*) Laboratoires d'Electronique et de Physique appliquée, Paris.

[1]) V. V. Schwab and J. G. van Wijngaarden, The EC 59, a transmitting triode with 10 W output at 4000 Mc/s, Philips tech. Rev. **20**, 225-233, 1958/59 (No. 8). In some earlier publications the EC 59 is referred to as 49 AF. — The tube EC 157 differs from the older type EC 57 in that its cathode has a longer life.

[2]) J. P. M. Gieles, A 4000 Mc/s wide-band amplifier using a disc-seal triode, Philips tech. Rev. **19**, 145-156, 1957/58; referred to henceforth as I.

The value of L_k in the EC 157, however, is approximately 3×10^{-9} H, whereas in the EC 59 it is only 10^{-9} H. If we now replace the input impedance of the tubes by a resistance and a reactance in parallel,

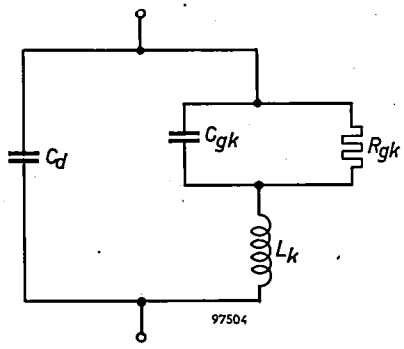


Fig. 2. Equivalent circuit for the input of a disc-seal triode. R_{gk} input resistance between grid and cathode; C_{gk} capacitance between these electrodes; L_k inductance of cathode holder and cathode bush; C_d capacitance between grid disc and cathode bush.

this resistance will be about 400 Ω in the case of the EC 157 at 4000 Mc/s, which roughly corresponds to the characteristic impedance of the waveguide. With the EC 59, on the other hand, this resistance is a mere 10 Ω. This makes it necessary to introduce an impedance transformer between the input waveguide and the tube.

For this purpose a quarter-wavelength transformer can be used. If this is made of a normal rectangular waveguide, however, the frequency band in which good matching is possible is found to be fairly narrow. A simple method of reducing this frequency-dependence is to use a ridge waveguide [3])

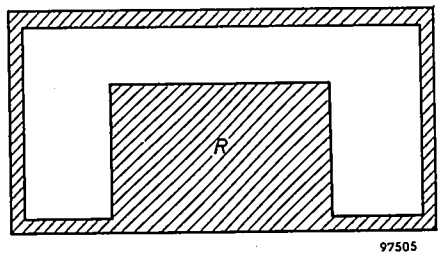


Fig. 3. Transversal section through a waveguide with ridge R, as used for the quarter-wavelength transformer for matching the input impedance of the tube to the characteristic impedance of the input waveguide.

having a transversal section as shown in *fig. 3*. Externally the dimensions are the same as those of the ordinary waveguide. The ridge considerably reduces the characteristic impedance of the waveguide, giving it properties resembling those of a parallel-wire transmission line. When a short section of waveguide (about $\frac{1}{4}\lambda$) in front of the cathode is provided with such a ridge of appropriate dimensions, good matching can be achieved in a frequency band almost twice as wide as when a normal quarter-wavelength transformer is used. The construction is illustrated in *fig. 4*. The tube is inserted from underneath and, the grid disc being threaded, screwed into the upper side of the waveguide. The cathode is connected to the lower side of the waveguide (in this case the top of the ridge) by means of a choke coupling (see I). The standing-wave ratio in the input waveguide of the amplifier equipped with an average tube is shown in *fig. 5* as a function of frequency (curve *a*). This curve is found when the anode circuit is detuned with respect to the reference frequencies so as to eliminate its effect on the input impedance of the tube.

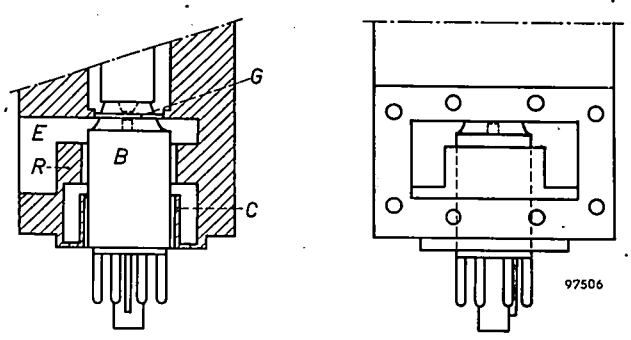


Fig. 4. Simplified representation of the input section of the amplifier. E input waveguide; B cathode bush; R ridge; C choke coupling; G grid disc.

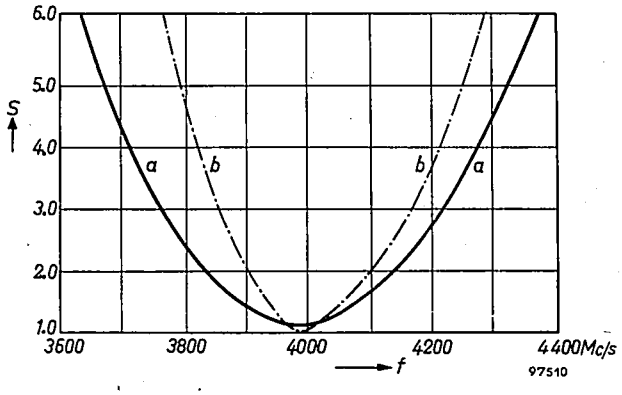


Fig. 5. Standing-wave ratio s in the input waveguide as a function of frequency f. Curve *a* applies before the reactive elements for matching the tube to the waveguide have been introduced. Curve *b* applies after their introduction, the tube being matched to the waveguide at 4000 Mc/s. In both cases the anode circuit is detuned with respect to the reference frequency.

[3]) See e.g. G. C. Southworth, Principles and applications of waveguide transmission, Van Nostrand, New York, 1950, p. 134 et seq., and S. B. Cohn, Properties of ridge waveguides, Proc. Inst. Rad. Engrs. **35**, 783-788, 1947.

If we take the bandwidth to mean the difference between the frequencies at which the load resistance of the waveguide receives half the maximum power, a simple calculation shows that this bandwidth is equal to the difference of the frequencies at which the standing-wave ratio is $3 + 2\sqrt{2} = 5.8$. From fig. 5 we accordingly find a bandwidth of 700 Mc/s.

In order to be able to match individual tubes accurately to the waveguide at any desired frequency in the band to be covered, two variable reactive elements are introduced into the waveguide as in I. The first is located near the beginning of the ridge, and the second at about an eighth of a wavelength before the ridge. These elements are not shown in fig. 4; they can, however, be seen in the photograph of the complete amplifier in *fig. 6*.

In fig. 5, curve *b* represents the standing-wave ratio in the input waveguide when the above matching elements are fitted and when the tube (again with detuned anode circuit) is matched to the input waveguide at a frequency of 4000 Mc/s. We see that the introduction of the matching elements has reduced the bandwidth to about 500 Mc/s.

When the tube is switched off the input impedance is obviously different and the tube is no longer matched to the waveguide. In this case the standing-wave ratio in the input waveguide is much larger, being about 20 [4]).

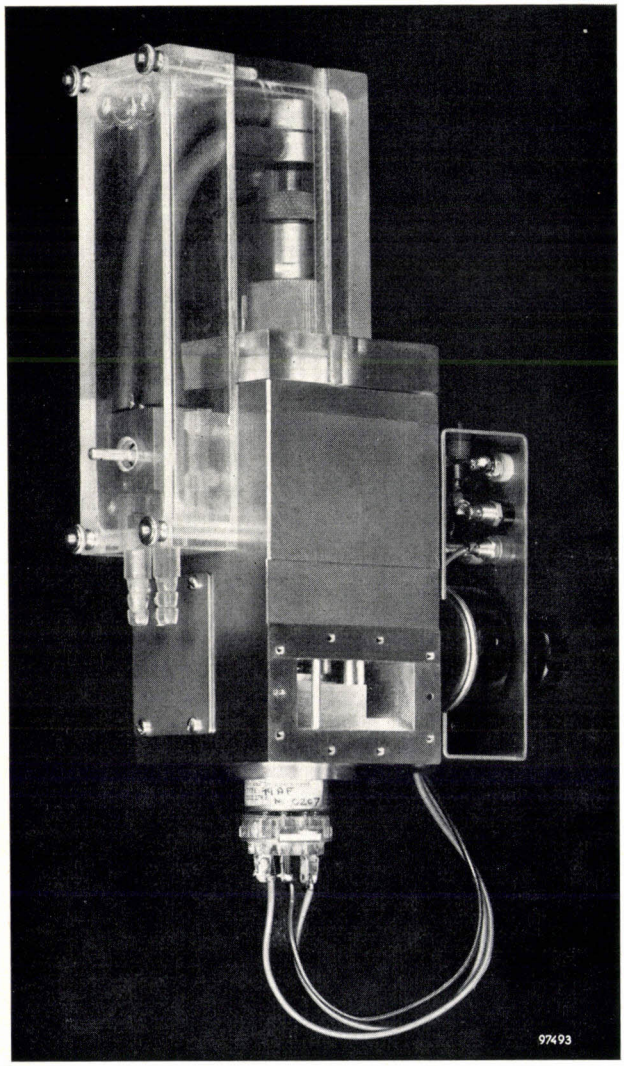


Fig. 6. Amplifier with EC 59 triode, seen from the input side. The safety cover on the left contains, among other things, the anti-corrosion block. In the input waveguide the two reactive elements for matching the tube to the waveguide can be seen.

The output circuit

As in the amplifier for the EC 157, the anode of the tube is contained in a resonant cavity which is terminated at the top by a plunger. The anode is again provided with an extension piece constituting the inner conductor of a coaxial transmission system. The first section of this system, i.e. the part inside the plunger, acts as a quarter-wavelength transformer. The high-frequency energy passes between the plunger and this inner conductor, and arrives in the second section of the coaxial line. This line must be properly matched to the output waveguide. The design of the transition from this line to the output waveguide is complicated in this case by the fact that the inner conductor, the lower end of which is joined to the anode of the tube, has to contain the tubes for supplying and removing the cooling water. This makes it necessary that the inner conductor should reach the opposite wall of the waveguide, so that a probe-construction as in the EC 157 amplifier is not possible. If the inner conductor makes high-frequency contact with the opposite wall of the waveguide, the characteristic impedances of coaxial line and waveguide must be equal. The characteristic impedance of the waveguide is approximately 500 Ω, but of the coaxial line only about 20 Ω, the reason being that the dimensions of the coaxial line are governed by those of the tube and resonant cavity. Here too, therefore, impedance transformation is needed. The use of a single quarter-wavelength transformer was not sufficient in this case, since the frequency range in which a matched transition is possible is too small. It was found that a transition from coaxial line to waveguide requiring no means of correction in the entire frequency band from 3800 to 4200 Mc/s could be achieved with two quarter-wavelength transformers, one in the coaxial line and the other in the waveguide.

[4]) This confirms that the input impedance of the tube is closely dependent on the transconductance, and is not caused, for example, by losses in the materials.

This is illustrated in *fig.* 7. The standing-wave ratio of this transition is smaller than 1.2 within the frequency band mentioned, and the space inside the narrowed section of the inner conductor is sufficient to accommodate cooling-water tubes.

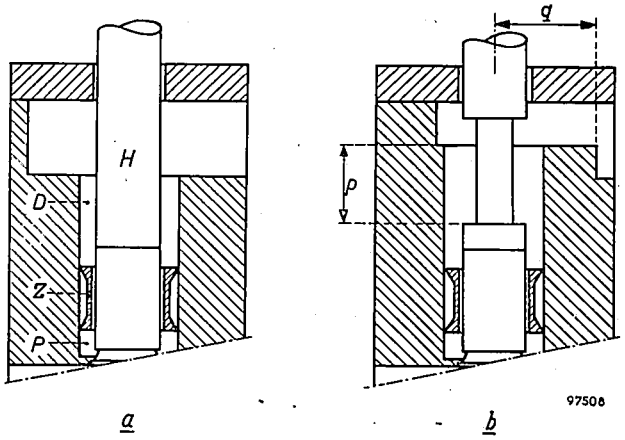


Fig. 7. Simplified representation of the output section of the amplifier, *a*) without quarter-wavelength transformers, *b*) with two quarter-wavelength transformers, one being included in the coaxial output line and the other in the output waveguide. *D* cylindrical hole constituting, together with the inner conductor *H*, the coaxial output line. *Z* plunger for tuning the anode resonant cavity *P*. The distances denoted by *p* and *q* are approximately a quarter wavelength.

The upper part of the inner conductor must make high-frequency contact with the waveguide and nevertheless be insulated from it for direct voltage. This is again achieved with a choke coupling. In view of the high voltage (500 V) and the narrow gap (0.5 mm) it is important that the inner conductor be very accurately centred. Since it is screwed to the anode, and the tube is fixed with respect to the amplifier block by the grid disc, only slight eccentricity of the anode would be enough to cause a short-circuit. For this reason a flexible section is incorporated in the inner conductor at the position where the constriction begins (denoted by *B* in fig. 8). The wall thickness of the horizontal face at that point is only 0.3 mm, making it possible to move the upper part slightly with respect to the lower, and thus to centre the upper part with respect to the amplifier block. If this were done automatically when screwing-in the tube, there would be a risk of damaging the grid screw thread. The construction is therefore such that the tube, fitted with the inner conducter, is first freely screwed in, after which a centering cap is attached to the lid. This cap is also utilized for leading-in the high tension. As a safety precaution the cap is completely surrounded by an insulating cover (see fig. 6).

The amplifier, as hitherto described, will show a frequency characteristic with a single peak and have a bandwidth of 100 Mc/s. As in the case of the amplifier using the EC 157, however, we require here too a bandpass-filter curve with transitional coupling. If a second resonant circuit is introduced in the output waveguide for this purpose (an "iris", see I), it must be so positioned that the system between the top of the plunger and the iris is electrically equivalent to an uneven number of quarter wavelengths. In the amplifier for the EC 157 this distance was equivalent to a line one quarter wavelength long. Owing to the two quarter-wavelength transformers required in the output circuit for the EC 59, this distance could not be maintained. The distance between the iris and the top of the plunger in this case therefore had to be electrically equivalent to three quarter-wavelengths. It was found, however, that with one fixed place for this iris it was no longer possible to satisfy the above-mentioned condition with sufficient accurary for all frequencies in the range from 3800 to 4200 Mc/s. For this reason the iris for the EC 59 is not mounted in the amplifier, but screwed to the output flange as a separate unit. The frequency band to be covered is now divided into two parts, one from 3800 to 4000 Mc/s and the other from 4000 to 4200 Mc/s. For each part a separate unit is used, having a different effective iris-to-plunger distance. Each unit also contains a screw used for adjusting the coupling to the required value (see I).

When the anode circuit is now tuned and adjusted to constitute a band-pass filter with transitional coupling, the bandwidth between the points where the gain has dropped 0.1 dB with respect to the gain at the central frequency is found to be 55 Mc/s, as in the case of the EC 157 amplifier. Since the signals to be transmitted cover a frequency band of 20 Mc/s, this is more than adequate.

The cooling system

The anode of the EC 59 dissipates 125 W (500 V, 250 mA). This power is developed on the anode surface, the area of which is about 12 mm^2. In order to keep heat generation on the surface within reasonable bounds, water cooling is employed. Since a power of 300 W can be continuously dissipated by a stream of water of 0.5 litres per minute, this rate of flow is amply sufficient.

The anode is hollow and internally threaded. The inner conductor of the coaxial output system, is screwed into the anode and is provided with a bore intended for carrying off the cooling water, which is introduced through a thin pipe passing centrally through the bore and terminating just

above the actual anode surface. A sketch of the arrangement is shown in *fig. 8*. A watertight seal is provided by a rubber ring. The contact between the metal parts is external, so that the rubber ring does not affect the electrical performance.

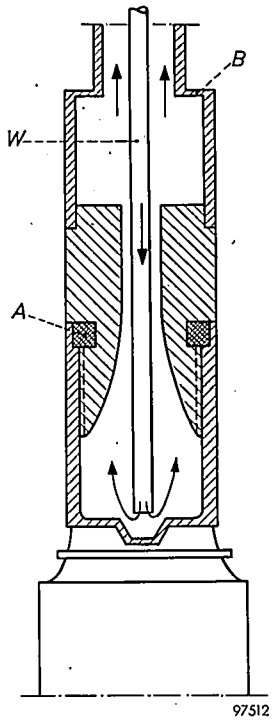


Fig. 8. Simplified cross-section of the inner conductor of the coaxial output line. *W* feed pipe for cooling water; *B* flexible portion to allow centering in the block; *A* rubber ring. The arrows indicate the water flow.

After the tube has been introduced into the amplifier block, a special union which converts the coaxial water supply into a twin-duct system is fitted to the top of the inner conductor. The supply duct and the outflow duct pass an insulated metal block attached to the side of the amplifier and kept at the same potential as the anode. The idea underlying this anti-corrosion block, as it is termed, is the following. The cooling water always contains impurities (even in a closed water-circuit), which give rise to electrolysis and hence to corrosion. The process is accelerated by the fact that the anode material corrodes easily and is difficult to provide with a protective plating. Since the first portion of the part under high tension is the most severely attacked, the water is first conducted through the metal block mentioned and then to the amplifier proper. As a result, this block corrodes first, and can later be replaced, whilst the amplifier itself is not affected. Since the anti-corrosion block is under high tension, it is also included under the safety cover (see fig. 6).

Properties of the complete amplifier

Small signals

The gain obtained with most of the tubes tested is about 10 dB in the case of small signals at a bandwidth of 100 Mc/s. If, after adjusting the anode circuit as described above, we again measure the standing-wave ratio in the input waveguide, we find it to be approximately 1.5 at the central frequency. Here, then, the feedback is less than with the EC 157, for which the corresponding value is about 2. The reason is to be found in the entirely different configuration in the effective triode section. The frequency at which no internal feedback occurs is about 5200 Mc/s with the EC 157, whereas with the EC 59 it lies in the region of 4200 Mc/s. This is much closer to the operating frequency, which explains the smaller feedback.

Strong signals

Fig. 9 shows a plot of the gain versus output power. The curve represents the average of measurements made on 16 tubes. The anode voltage was 500 V and the anode current 250 mA. The amplifier was again adjusted for a small-signal bandwidth of 100 Mc/s, at which the gain was almost 10 dB. At an output of 10 W the average gain was 8.6 dB [5]). The gain was still found to be more than 6.5 dB at an output power of 20 W, which would call for an input power of 4 W. If a tube is to be loaded continuously with such a high input power, however, measures must be taken to prevent overloading of the grid. It has been found that the maximum possible driving power is approximately reached at a grid current of 15 mA. This grid current depends among other

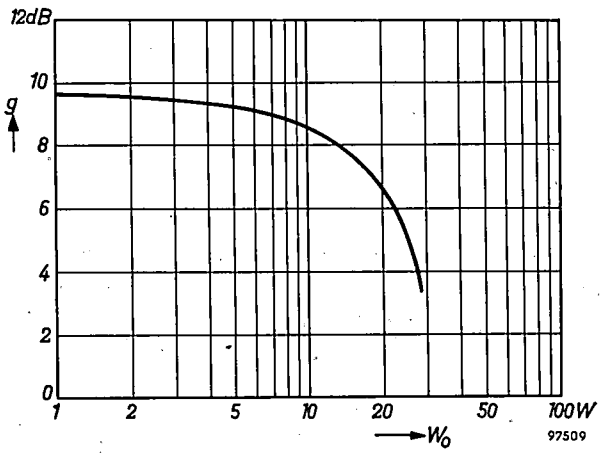


Fig. 9. Gain *g* as a function of power output W_0 of the EC 59 amplifier. Bandwidth of the amplifier 100 Mc/s. Anode voltage 500 V, anode current 250 mA.

[5]) The guaranteed value is 7 dB.

things on the static I_a-V_g characteristic of the tube and thus differs from tube to tube. With most tubes we were able to produce an output of 15 W without exceeding this limit. The gain was then 7 dB and the input power 3 W [6]). If a higher output is required, it can be achieved by increasing the anode voltage whilst keeping the anode current constant. This has little effect on the curve in fig. 9, but the negative grid voltage is now higher with the same anode current. The grid current therefore drops, thus permitting a higher driving power.

The driving power for the EC 59 can be obtained in various ways. One way is to use an EC 157 for the purpose. This delivers 1.5 W, which is amplified by the EC 59 to more than 10 W. One can also use two EC 157 amplifiers connected in parallel, which can readily be done with microwave triode amplifiers by means of power dividers in the waveguides. In this way twice the power is produced for the same gain, so that with two EC 157 tubes an output of 3 W can be obtained. This is amplified by an EC 59 amplifier to 15 W.

If an even higher power output is required, there is the further possibility of connecting two EC 59 amplifiers in parallel. With two parallel EC 157 amplifiers as the driving stage, an output of more than 20 W can be achieved. Driving the double output stage with a single EC 59, in its turn driven by an EC 157, can give 30 W without the maximum permissible grid current being exceeded.

In all these circuits the cascaded amplifier stages should be coupled by means of directional isolators. As explained in I, this considerably reduces many difficulties arising from feedback.

Group delay, AM-PM conversion

In the foregoing we have discussed the properties of the amplifier only with regard to the amplitude of the signals. In radio links using frequency-modulation it is also important, however, to examine the phase shift between the input and output signals. We can measure this phase shift φ as a function of frequency and as a function of the output signal; from the first-mentioned relation we can calculate the group delay, $d\varphi/d\omega$. We can also measure the group delay directly.

Extensive group-delay measurements were carried out at the time on the EC 157 amplifier [7]). One of the conclusions was that the group delay of a triode amplifier is determined almost entirely by the circuit employed, and is not noticeably affected by the tube or the signal level. Since it could safely be assumed that this would also apply to the EC 59 amplifier, which after all has the same bandwidth, no group-delay measurements were carried out in this case. The group delays, including the variations in group delay within the frequency band of the transmitted signals, are expected to be the same as those occurring in the EC 157 amplifier.

The situation is different, however, when we turn from the group delay $d\varphi/d\omega$ to the phase shift φ, for this does depend on the signal level. Since we are concerned here with frequency modulation, all the signals involved might be assumed to have a constant amplitude, in which case the above consideration would be of no importance. In practice, however, a constant signal level is seldom possible. As a rule there is also some slight amplitude modulation of the input signal, giving rise to a phase variation which, in frequency modulation, causes distortion. This AM-PM conversion, as it is termed, is expressed in degrees of phase variation of the output voltage resulting from a 1 dB variation of the input signal. The AM-PM conversion was measured on the EC 59 as a function of the output power. It was found to amount to 1.2° per dB at a power of 15 W. Compared with other microwave tubes, this is a particularly low value. With two amplifiers in parallel the AM-PM conversion remains the same, making it possible to obtain an output of 30 W with an AM-PM conversion of only 1.2° per dB.

[6]) Since sufficient life-tests have not yet been carried out at this power, users are recommended to keep the input power of the tube lower than 2 W for practical purposes.

[7]) See article I and also: J. P. M. Gieles, The measurement of group delay in triode amplifiers at 4000 Mc/s, L'Onde électrique **37**, 781-788, 1957.

Summary. An amplifier designed for use with the disc-seal triode EC 59 is described. The amplifier, which operates in the frequency band from 3800 to 4200 Mc/s, was evolved from an earlier-described type built for the EC 157 triode. Differences in the construction of the two tubes called for considerable modifications both at the input and output side of the amplifier block. The need for water cooling of the anode also involved constructional changes. Corrosion of the tube or amplifier block is prevented by conducting the cooling water through a detachable metal block kept at the same potential as the anode. At an output of 10 W and a bandwidth of 100 Mc/s the average gain measured on a series of tubes was 8.6 dB. An output of 15 W is obtainable without overloading the grid. If higher outputs are required, two EC 59 amplifiers can be connected in parallel.

As with the EC 157 amplifier, the group-delay variations in the frequency band concerned are deduced to be very small. Phase variations due to changes in the amplitude of the signal are also very small, being about 1.2° per dB.

EXPERIMENTS IN THE FIELD OF PARAMETRIC AMPLIFICATION

621.375.9

In many laboratories a great deal of work is now being put into developing the theory and practice of parametric amplification. The principle of this method of amplification is by no means new, but having long lain dormant it has come into prominence only in recent years as a result of research on the low-noise amplification of very weak signals as occur, for example, in radio astronomy and man-made satellite communications. In amplifiers equipped with the more or less traditional thermionic valves a fundamental limit is set to the reduction of the noise level by the random fluctuations of the electron emission from the hot cathode, and in amplifiers equipped with transistors by the statistical nature of the diffusion and recombination of charge carriers, which in this case determine the amplification. Such effects do not necessarily enter into parametric amplification, which is based on the periodic variation of one parameter of an oscillatory system, e.g. the stiffness in a mechanical system or the capacitance in an electrical circuit. In principle it should be possible with this method of amplification to achieve an extremely low noise level [1]).

Although numerous theoretical studies have been published on the mechanism of parametric amplification, and it has also been successfully applied in practice, it cannot be said that its operation is understood and under control to the same extent as the operation of amplifiers fitted with valves and transistors. For example, it is not yet possible to produce an optimum design for a parametric amplifier required to operate with a given frequency characteristic (or bandwidth). The interpretation of experiments in this field is often made more difficult by the absence of exactly defined conditions, so that various effects have to be dealt with that cannot be distinguished from each other.

Our own work on this subject has led to the construction of a simple experimental arrangement which permits the accurate testing of certain fundamental ideas and with which further experiments can be made. In this article we shall briefly describe this arrangement and some results of experiments. First of all, however, we shall explain summarily the principle of parametric amplification.

If a taut, slightly elastic string be pulled lengthwise at the appropriate frequency, the string will enter spontaneously into transverse vibration with a frequency equal to half that at which it is pulled (Melde's experiment, *fig. 1*). An LC resonant circuit in which the capacitance is periodically varied may be regarded as the electrical analogue of this

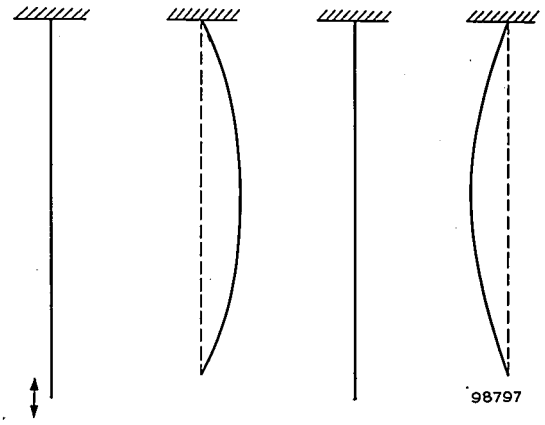


Fig. 1. When an elastic string is periodically pulled lengthwise with a frequency F, the string enters spontaneously into transverse vibration at its resonance frequency f, provided that $F = 2f$ (Melde's experiment, Ann. Physik u. Chemie **109**, 193, 1860).

long familiar phenomenon. Let us suppose that we have a parallel-plate capacitor, the distance between the plates being d_0, and that an alternating current flows through the (resistanceless) circuit at the resonant frequency $f = 1/(2\pi\sqrt{LC})$. The charge on the capacitor, and hence the voltage, varies sinusoidally with time. We now pull the two plates of the capacitor abruptly apart to a distance $d_0 + \delta$ (*fig. 2*) at the moment when the charge and thus

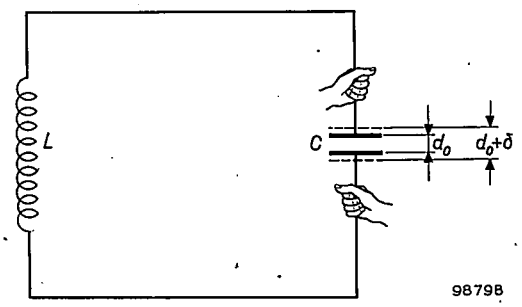


Fig. 2. LC resonant circuit in which the plates of the capacitor are pulled apart and pushed back again at certain moments.

the voltage of the capacitor reaches a maximum — irrespective of which of the plates is positive — and we push the plates back to their original position at the moments when the plates are uncharged and

[1]) As far as we know this was first suggested by A. van der Ziel, J. appl. Phys. **19**, 999, 1948. Electrical circuits with varying parameters were dealt with by B. van der Pol, Experimental Wireless **3**, 338, 1926.

there is thus no voltage difference between them. The first action requires work, the second produces no work, and so energy is periodically supplied to the circuit. The result is that the amplitude of the alternating voltage increases, i.e. "negative damping" occurs. The process is further elucidated in *fig. 3*. Even though no current flows at the start, a

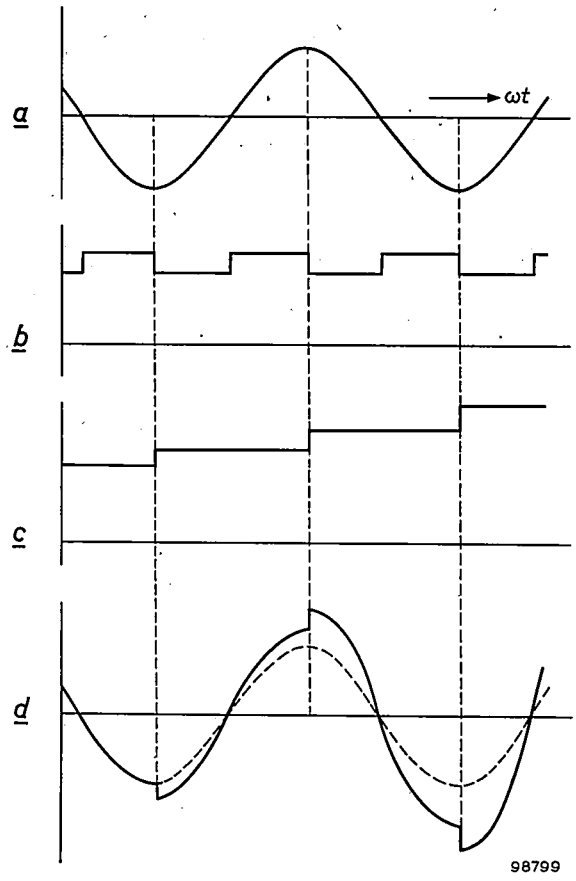


Fig. 3. Mechanism of parametric amplification in the circuit of fig. 2. In the initial state a small alternating current flows in the (resistanceless) circuit. The charge and voltage on the capacitor then vary sinusoidally with time (*a*). The capacitor plates are now displaced at certain moments so as to vary the capacitance as shown by the square wave form in (*b*). This "pumping" causes a stepwise increase of the energy in the circuit (*c*), and also of the amplitude of the voltage on the capacitor (curve *d*).

minute disturbance of equilibrium in the form of an extremely small charge — which is always present — is sufficient to initiate the process, and this causes the LC circuit to enter spontaneously into electrical oscillation. The frequency at which the capacitance is varied (by moving the plates back and forth), called the *pump frequency*, is most favourable when it is twice the resonance frequency of the circuit — just as in the mechanical case.

The negative damping of a circuit, which may be regarded as an effect arising from a negative resistance, implies the possibility of using it to amplify a given input signal. In this respect, however, the situation is not so simple, for one thing because of the frequency condition just referred to. The difficulty is made clear by the following consideration. For negatively damping an oscillation it is not necessary that the pumping should be exactly in phase with the oscillation (*fig. 4*). The net energy supply from the pump falls, however, as the phase angle φ increases. The supply drops to zero at $\varphi = 45°$, and becomes negative at phase angles larger than this; in other words, the oscillation is then damped instead of amplified. The damping is greatest at $\varphi = 90°$. Now if half the pump frequency (i.e. the resonance frequency f of the circuit) differs slightly from the frequency of the input signal, this is equivalent to a continuous change of phase; the oscillation is thus alternately amplified and (though to a lesser degree) damped; in other words, it is modulated in amplitude.

In attempts to arrive at a practical means of parametrically amplifying radio signals, where the aim is to achieve useful amplification in a reasonably wide frequency band, the difficulty described has been circumvented by the use of more complicated

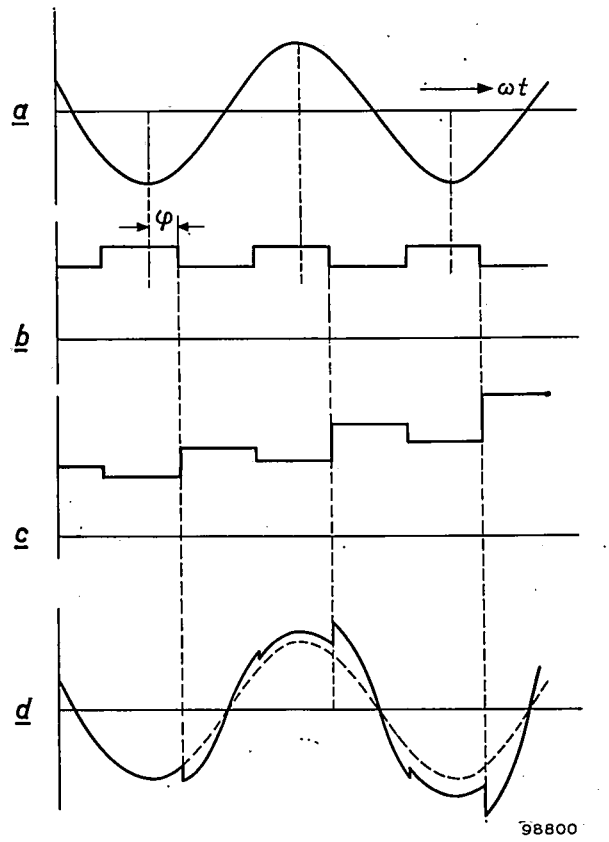


Fig. 4. If the capacitance variation (*b*) is not in phase with the alternation of the charge on the capacitor (*a*), but differs by a phase angle φ, the energy in the circuit alternately rises a step and drops a step (*c*). The steps up are greater than the steps down when $0 < \varphi < 45°$ or $135° < \varphi < 180°$. The capacitor voltage thus increases (*d*); at $45° < \varphi < 135°$ it decreases.

circuits with two or more resonance frequencies [2]). Another question concerns the method of carrying out the pumping process. The abrupt capacitance variation, effected by moving the plates to and fro by hand as in fig. 2, is obviously not intended for practical purposes. But any other mechanical variation of an oscillatory-circuit parameter is equally unpracticable because of the excessive acceleration forces involved at the very high frequencies with which we are concerned. In the attempts above mentioned, the variation is therefore effected by means of *voltage-dependent capacitances* or current-dependent inductances. An example of a voltage-dependent capacitance is the germanium diode, which can be used in the centimetric wave range. Its capacitance depends on the negative bias. If we superimpose on this an alternating voltage (the pump or control voltage), we then have a periodically varying capacitance. Parametric amplifiers have already been built in this way, and are reported to possess, as predicted, favourable noise properties [3]). These systems, however, are difficult to deal with theoretically, for the capacitance of the diode depends not only on the pump voltage but also on the oscillatory-circuit voltage to be amplified. (A further complication is the non-linear relation between capacitance and voltage.) Such systems are called *autoparametric* as opposed to *heteroparametric* systems (as in fig. 2), in which the parameter to be varied is influenced only by the pump, and does not react noticeably on the pump. If a current-dependent inductance is used, e.g. a coil or resonant cavity having an iron or ferroxcube core which is alternately saturated by a pump current (essentially a "magnetic amplifier" long used for low frequencies), this, too, is an autoparametric system, since the circuit current itself likewise contributes to the saturation.

Referring to fig. 2, if the plate distance varies according to $d = d_0 + \delta \cos 2\omega t$, i.e. the capacitance according to

$$C = \frac{C_0}{1 + \frac{\delta}{d_0} \cos 2\omega t}, \quad \ldots\ldots \quad (1)$$

then the charge q of the capacitor in this heteroparametric system is given by Mathieu's equation (linear second-order differential equation with time-dependent coefficients):

$$\frac{d^2q}{dt^2} + \frac{1}{LC_0}\left(1 + \frac{\delta}{d_0} \cos 2\omega t\right) q = 0. \quad \ldots\ldots \quad (2)$$

The solutions of this equation, one of them with the angular frequency $\omega = 1/\sqrt{LC_0}$ (i.e. half the pumping frequency), are known and their properties have been extensively investigated [4]). They can thus be manipulated, even though the treatment is not simple.

In autoparametric systems, on the other hand, the differential equations involved are non-linear, and in general cannot be solved exactly but only by methods of approximation. It is then very difficult if not impossible to obtain any general idea of the properties of the solutions.

Thus, although practice points in the direction of autoparametric systems, it is nevertheless desirable for the purposes of fundamental research to confine oneself in the first place to the heteroparametric systems, such as the mechanically variable capacitor, which lend themselves more readily to theoretical treatment.

In the thirties attempts had already been made to vary the capacitance mechanically by using a motor-driven rotating capacitor as the pump [5]). It was in fact possible in this way to demonstrate the spontaneous onset of oscillations in an LC circuit, having a resonance frequency of 28 kc/s.

The experimental arrangement which we have devised is of a similar kind, but the capacitance variation is effected by a more manageable element, namely a magnetostrictive resonator, which also offers more possibilities for further experiments. The pump consists of a window-frame core, shown in *fig. 5*, of a suitable type of ferroxcube, widely used nowadays for ultrasonic oscillators [6]). This core is centrally clamped and is brought into longitudinal vibration by means of a winding connected to an ultrasonic generator of frequency 22 kc/s. To each of the two parallel-ground end faces a metal plate of 30×80 mm is attached. Opposite each plate, at a spacing $d_0 = 80$ μ, a parallel, fixed electrode is mounted. The two moving plates are electrically interconnected and constitute one plate of a capacitor, whilst the fixed electrodes, also interconnected, constitute the other plate. As can easily be calculated, the total capacitance amounts to 532 pF. To obtain an oscillatory circuit having a resonance frequency $\omega/2\pi$ of 11 kc/s (i.e. half the pump

[2]) The phase-dependence of the negative damping is usefully exploited in the "Parametron" introduced by Japanese investigators. In this device the two states of either oscillating or not are used for registering one bit of information, in a manner similar to the use of the two states of magnetization of a ferrite core.

[3]) A. Uhlir, Junction-diode amplifiers, Scientific American **200**, No. 6, 118 et seq., June 1959. — A review of parametric amplification, together with an extensive bibiography, can be found in H. Heffner, Solid-state microwave amplifiers, Inst. Rad. Engrs. Trans. **MTT-7**, 83-91, 1959 (No. 1).

[4]) See e.g. E. T. Whittaker and G. N. Watson, A course of modern analysis, Cambridge University Press, London 1920, third edition, Chapter 19, § 19.7.

[5]) L. Mandelstam, N. Papalexi, A. Andronov, S. Chaikin and A. Witt, Exposé des recherches récentes sur les oscillations non linéaires, Tech. Phys. USSR **2**, 125-127, 1935.

[6]) See C. M. van der Burgt, Ferroxcube material for piezomagnetic vibrators, Philips tech. Rev. **18**, 285-298, 1956/57.

frequency, see above), an inductance L of 0.395 henry is needed.

One of the advantages of this arrangement is that it is easy to produce with it a purely *sinusoidal* variation of the reciprocal capacitance. In that case an exact theoretical treatment is possible (we have the case of eq. (1) and (2), the Mathieu equation).

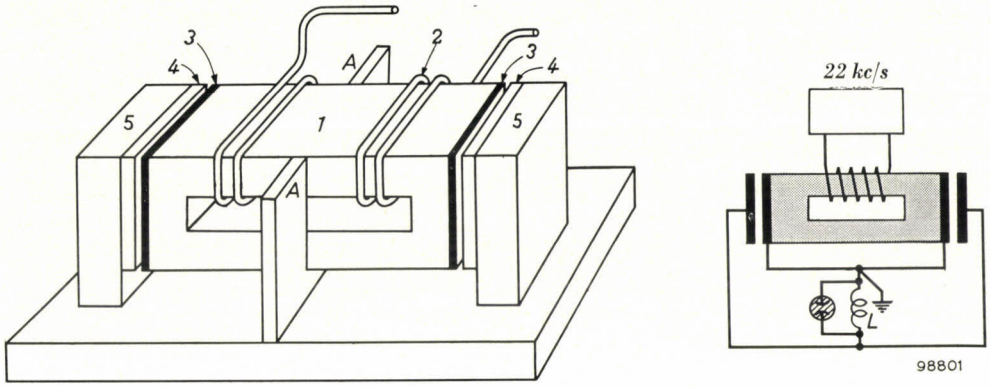


Fig. 5. Construction of a "magnetostrictive pump". *1* frame core of piezo-magnetic ferroxcube [6]), clamped at *A*; *2* core winding through which the current from an ultrasonic generator flows at a frequency of 22 kc/s; *3* metal plates attached to the core; *4* fixed plates spaced 80 μ from *3*; *5* insulating material. The basic diagram of the arrangement is shown on the right. The coil L must meet certain requirements regarding its inductance and its Q.

In particular it can be demonstrated that the parametric excitation of the natural frequency of the circuit calls for a certain minimum amplitude δ of the capacitor plates, and that this minimum amplitude is smaller the higher the Q of the resonant circuit ($Q = \omega L/R$, R being the series resistance of the coil). Provided δ/d_0 is sufficiently small, we find the condition (corresponding to the limit of the first stability region of the Mathieu equation [4])):

$$\frac{\delta}{d_0} Q > 2. \quad \ldots\ldots \quad (3)$$

With our ultrasonic oscillator the breaking strength of the core limits the amplitude δ to 1 micron. It follows from equation (3) that the coil in the resonant circuit must accordingly be constructed such that the Q of the circuit is at least 160.

The experiments confirm that this condition must indeed be fulfilled. *Fig. 6* shows two coils which both have the required inductance and more than a sufficient Q, one with an air core and the other with a ferroxcube core.

The amplitude reached by the voltage on the capacitor when the oscillation increases due to the pumping depends, in theory, solely upon the pump power. The power supplied to the circuit is $P = \frac{1}{2}\omega C_0 V^2/Q$. At $C_0 = 532$ pF and $Q = 200$, a voltage as high as $V = 3300$ V is reached at a power of only 1 W, which constitutes a scarcely noticeable load on the magnetostrictive resonator. A voltage as high as this would obviously cause our capacitor, with its air gap of 80 μ, to break down. Using the air-core coil in our arrangement it was indeed found necessary to limit the voltage in order to protect the capacitor. This was done by shunting a neon tube across the circuit, which is visible in the photograph of the complete set-up shown in *fig. 7*. This precaution is not required when the coil with

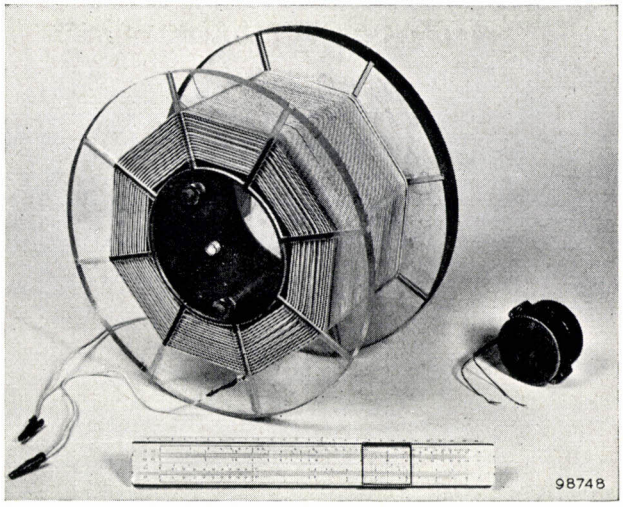


Fig. 6. Two coils used for demonstrating parametric amplification by means of the magnetostrictive pump. Left, coil with air core; right, coil with ferroxcube core.

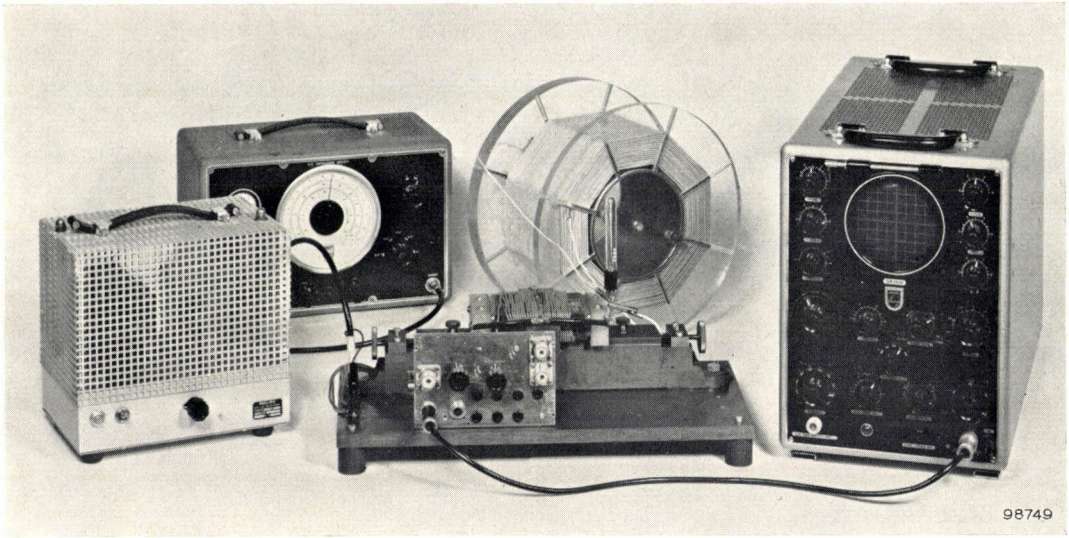


Fig. 7. The complete set-up. In the middle the magnetostrictive resonator with the air-core coil behind it (cf. fig. 6). A neon tube is shunted across the latter to limit the voltage. Left, the ultrasonic generator consisting of a signal generator and amplifier; right, oscilloscope for observing the amplified oscillation.

ferroxcube core is used: in this case the capacitor voltage does not rise above 250 V, owing to saturation of the ferroxcube.

The set-up illustrated can also be used as a simple means of demonstrating parametric amplification. The circuit is shown in *fig. 8*. The pump amplitude is set just below the threshold of spontaneous oscillation. As input signal we can take, for example, a signal of 11 kc/s, applied in the right phase from a signal generator. The signal can then be measured

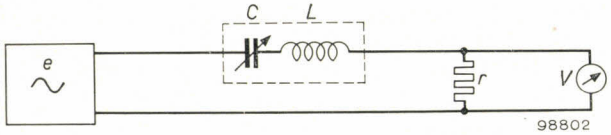


Fig. 8. Circuit for demonstrating parametric amplification with the set-up described; *e* signal source (signal generator), *r* load resistance, *V* voltmeter.

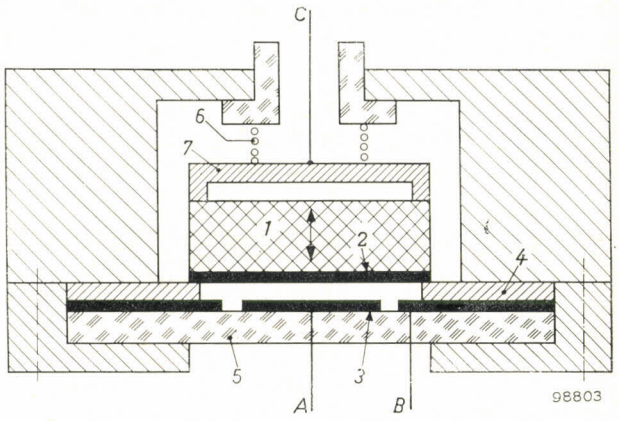


Fig. 9. Piezo-electric resonator as pump; pumping frequency 3 Mc/s. *1* resonating quartz crystal $19 \times 19 \times 0.95$ mm (X-cut), excited into the thickness compressional mode; *2* and *3* vapour-deposited silver layers; *4* metal spacing-ring 25 μ thick; *5* quartz-glass plate; *6* compression spring and *7* pressure plate; *A* and *B* connections for the inductance; *C* connection for exciting the quartz crystal.

at the output with the pump switched on and off. In this way we found a voltage amplification of 20 times.

We have made an arrangement similar to the one described above using a *piezo-electric* resonator as pump. The resonator consists of a quartz-crystal plate about 1 mm thick and with a surface area of about 20×20 mm^2, which is excited into the thickness compressional mode (*fig. 9*); the pump frequency is 3 Mc/s. The results obtained were similar to those obtained with the magnetostrictive pump.

B. BOLLÉE and G. de VRIES.

MODERN ACOUSTICAL ENGINEERING

II. ELECTRO-ACOUSTICAL INSTALLATIONS IN LARGE THEATRES

by D. KLEIS.

534.84:534.86:621.395.623.8

The various kinds of electro-acoustical equipment now available were surveyed in Part I of this article. The principles underlying their employment in order to improve the intelligibility of the spoken word, and in order to give music more "body" in halls with unsatisfactory acoustics, were also dealt with. Mention was likewise made of electro-acoustical installations for controlling the acoustics of a hall or auditorium. In recent years such installations have been built in several large theatres and opera houses, including the Palais de Chaillot in Paris and the Scala Theatre in Milan. These and other installations will be discussed here. It will be shown that many kinds of electro-acoustical equipment are necessary in such theatres, over and above installations for controlling their acoustics, and that closed-circuit television may also prove to be of value.

Introduction

The endeavour of modern acoustical engineering is to reproduce sound naturally. This means paying proper attention to the properties of the human ear and the acoustical properties of the hall or auditorium. Part I of this article [1]) dealt with the principles underlying the employment of electro-acoustical equipment, and certain conditions were deduced which have to be satisfied if natural sound reproduction is to be achieved.

One of the essential points that emerged is that direct and indirect sound have different functions: direct sound gives the hearer an impression of the sources and their spatial arrangement, while indirect sound gives him an impression of the hall. The characteristic feature of installations for controlling acoustics was stated to be the separate production of direct and indirect sound, each kind being given the volume and spatial distribution best suited to its purpose. Direct sound is reproduced, often stereophonically, by means of directional loudspeakers, and indirect sound is relayed through a delay and reverberation device to widely distributed, diffusely radiating speakers. It was further discussed in Part I how these methods might be employed (*a*) for obtaining the best possible reproduction of music and speech in halls and in the open air, (*b*) for supporting a live performance, by adapting the hall acoustics electro-acoustically to the type of performance being given, and (*c*) for introducing sound effects. At the same time mention was made of some general aspects of microphone positioning.

In Part II we shall start by studying the acoustical problems of theatres in general, go on to investigate what electro-acoustical facilities theatres require, and then discuss certain large installations [2]) which have been built in the following theatres and halls:

Teatro alla Scala in Milan,
Théâtre National Populaire (Palais de Chaillot) in Paris,
Grand Auditorium at the 1958 Brussels World Fair,
Gebouw voor Kunsten en Wetenschappen in The Hague.

Finally, a large public-address installation [3]), namely that at the Volkswagen factory at Wolfsburg, will be discussed.

Acoustical problems in theatres, and their solution

In Part I we saw that speech and various kinds of music make conflicting demands on the acoustics of a hall. That would not matter if a hall with suitable acoustics was available for each and every kind of performance. This is far from being the case even in many large towns; moreover, where suitable halls do exist, they often do not possess the technical facilities that a performance before a large audience requires. It is therefore inevitable that one and the

[1]) D. Kleis, Modern acoustical engineering, I. General principles, Philips tech. Rev. **20**, 309-326, 1958/59 (No. 11).

[2]) Designed and installed by the Projects Department of the Electro-Acoustics (E.L.A.) Division at Eindhoven, in consultation with the Acoustical Advisory Bureau and in collaboration with Philips organizations in the countries concerned and with technicians of the Commercial Department of E.L.A. Eindhoven.

[3]) Designed and installed by Deutsche Philips G.m.b.H.

same hall should have to serve for performances and gatherings of greatly differing kinds — symphony concerts, chamber music, opera, stage plays, meetings etc. Even if the question of acoustics be left aside, a concert hall is unsuitable for stage or opera performances because of lack of stage machinery, property rooms, stage lighting, dressing rooms and so on. However, reasons of this kind do not militate against the giving of concerts in a well-equipped theatre; it is only the acoustics of the stage and auditorium that make it undesirable to do so. In circumstances where one building has to suffice, then, it is logical to give it the full technical equipment of a theatre and to install electro-acoustical devices that make the auditorium equally suitable for musical performances.

It was demonstrated in the Grand Auditorium at the 1958 Brussels Exhibition that it is perfectly practicable and artistically permissible to put on performances of a widely differing character in a theatre that has been properly equipped from the electro-acoustical standpoint. During the six months of the Exhibition, the Auditorium, with its seating capacity of 2300 (*fig. 1*), was the scene of a great variety of meetings and shows, these taking place almost every evening and often during the afternoon as well; there were official addresses (including those at the ceremonial opening of the Exhibition), international congresses (with simultaneous interpreting), stage performances, operas, ballets, folk-dance programmes, song recitals, soloist and symphony concerts, choral performances, a film festival and a festival of electronic music. Thanks to an elaborate electro-acoustical installation it was possible for all this to take place to the satisfaction of both performers and the public.

It must not be concluded from the foregoing that the only reason for providing a theatre with an electro-acoustical installation is to make it suitable for concerts. More often the work is carried out in the first place with stage performances in mind. Provided it has been correctly designed and built, the installation serves the further purpose of making the theatre suitable for musical performances. In this way the building acquires a new usefulness.

Intelligibility

The biggest problem in many theatres is intelligibility. Imperfect intelligibility may be caused by an excessively long reverberation time, but it may also be due to the dimensions of the auditorium. Many modern theatres seat 1500 to 3000 people. In an auditorium of this size the sound may be so heavily attenuated, passing over the audience on its way to the back rows, that people seated there cannot hear it properly above the noise inevitably made by other members of the audience. The players have to make a disproportionate effort just to be understood, and this usually detracts from their performance without being entirely effective.

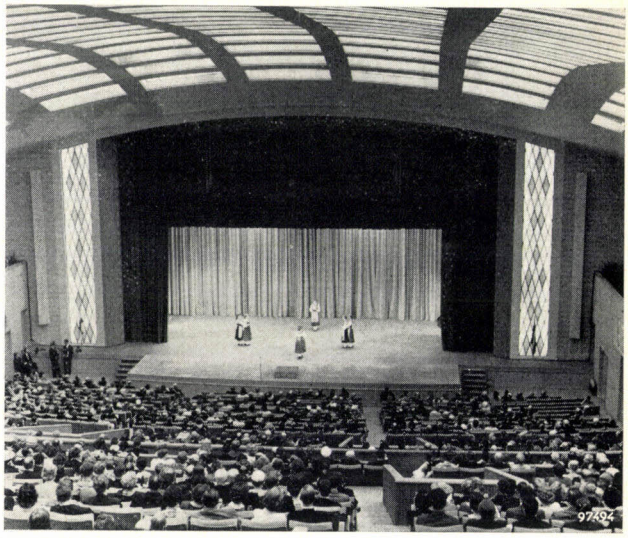


Fig. 1. The Grand Auditorium at the 1958 Brussels World Fair. The hall has a length of about 45 m, a cubic capacity of 15 000 m^3 and a seating capacity of 2300.

Irrespective of whether excessive reverberation or the size of the auditorium is the cause, poor intelligibility can be remedied by supplying additional direct sound from loudspeakers mounted on either side of the stage, and aiming these at the places where the stage proceedings are not being properly heard. As was made clear in Part I in the section headed *Direct sound*, sound from loudspeakers has the advantage, compared with the speaker's natural voice, of being beamed. What is more, if the loudspeakers are mounted high enough, the sound does not pass so closely over the heads of the people in the front of the auditorium and consequently suffers less absorption.

Where good intelligibility can be achieved with a loudspeaker signal not more than 10 dB louder than the direct sound from the actors on the stage, natural directional hearing can be maintained by introducing a slight delay into the signal (see Part I, p. 314). Sometimes the required delay can be obtained simply by mounting the loudspeakers in a particular position. However, if intelligibility necessitates raising the level of the loudspeaker signal more than 10 dB above that of the original direct sound, then stereophonic reproduction will be necessary if the audience are nevertheless to hear the stage proceedings from the right direction. The dangers of turning up the volume control too far

(and unhappily this is done all too often) were pointed out in Part I.

Though the direct sound may have to be amplified through 10 dB and sometimes even more, this does not mean that its loudness is increased in the same proportion. We shall demonstrate the truth of this in two cases, one for each cause of poor intelligibility.

In a large auditorium that nevertheless has a short reverberation time, high-pitched sounds are too weak on arrival at the rear, while low-pitched ones generally retain enough of their volume. Hence all the loudspeakers have to do is to reinforce the high-pitched sounds and so to restore the balance between these and the low-pitched ones. Now, high-pitched sounds contribute a great deal to intelligibility but very little to loudness; when therefore the loudspeakers are switched on or off, the difference in intelligibility is striking, though the difference in loudness is scarcely perceptible.

In an auditorium having too long a reverberation time the intensities of the direct and indirect sound are badly out of proportion for speech — say 1 : 20. If the intensity ratio is raised to 20 : 20 by means of an electro-acoustical installation, then the direct sound will be undergoing twenty-fold amplification, i.e. being amplified through 13 dB; but the total volume will only increase in the proportion of 21 : 40, i.e. by 3 dB. (Since it is mainly the high-pitched sounds that undergo amplification, the increase in overall loudness is still less.)

The foregoing will have made it clear that the term "amplifying equipment" is scarcely justified when applied to the type of installation described here, for there is hardly any question of an increase in loudness. On the other hand it is perfectly proper to use the term in connection with equipment for relaying sound in the open air, where the original sounds, both high-pitched and low-pitched, are too weak some distance away from the speaker.

There may be places in an auditorium — under a low balcony, for example — where the original sound does not penetrate sufficiently and where, often enough, the sound coming from the loudspeaker columns at the front of the auditorium is also too weak. Intelligibility in such places can be improved by fitting loudspeakers on the spot. The sound they give must be suitably delayed, to prevent the hearers getting the impression that it is coming from somewhere in their vicinity. (Here again there is usually no question of true amplification.)

Music

Music is an essential element in operas, operettas, revues, shows and also in many stage plays. Having dealt with intelligibility as the primary acoustical problem in theatres, we would reserve second place for the problem of making music sound as it should. Often, where intelligibility is good, music will not sound entirely satisfactory, the reason being that music requires a longer reverberation time than is compatible with good intelligibility. Added to this is the fact that the interior architecture of a theatre must satisfy certain visual requirements and these sometimes conflict with acoustical ones.

The orchestra providing the music for an opera or a play must not impede the audience's view of the stage. Accordingly, it is relegated to the orchestra pit in front of the stage (*fig. 2*), and there most of its members generally have the proscenium above their heads. This location, especially the part of the pit that is covered over, is acoustically a very bad one. The rather feeble impression frequently made by the overture to an opera, even when the orchestra numbers sixty or more musicians, may be attributed to their unfavourable location. For that very reason the orchestra platform of the Radio City Theater in New York is hydraulically raised whenever the orchestra is giving a performance which is not an accompaniment to some stage spectacle.

A further visual requirement is that the view of the stage should be good from every seat in the theatre. It may be satisfied by the classical theatre interior, in which several balconies and tiers of boxes rise one above the other. The arrangement is not a bad one from the viewpoint of intelligibility either (provided at least that the auditorium is not so lofty that the upper balconies are excessively remote from the stage). Even so, the balconies and boxes absorb a great deal of sound, hindering the diffusion of music and making the reverberation time too short. Needless to say, this applies as much to voices from the stage as to the music provided by the orchestra.

The lack of diffuseness and reverberation can be remedied by picking up the music with directional microphones, delaying it, and reproducing it with delay and diffuse reverberation via loudspeakers distributed around the auditorium (see under *Indirect sound* in Part I). This is what has been referred to in the past as stereo-reverberation [4]). In the meantime the new term *ambiophony* has been introduced for this electro-acoustical technique [5]).

The music played in the orchestra pit can be made to sound better if it is backed up by directional loudspeakers placed on either side of the proscenium

[4]) R. Vermeulen, Stereo-reverberation, Philips tech. Rev. 17, 258-266, 1955/56. See also [1]).

[5]) See footnote [11]) in Part I of this article.

Fig. 2. The auditorium of the Scala theatre, Milan. The acoustics for music are favoured neither by the conventional siting of the orchestra in an orchestra pit, nor by the boxes (only four of the six tiers can be seen in the photograph).

arch — the same loudspeakers in fact that are used for improving intelligibility. By dint of such backing, the orchestra is placed "on stage" acoustically and so lifted out of the unfavourable location to which it is tied for visual reasons.

Reproduction by these loudspeakers ought to be stereophonic in order to make the music "transparent" (i.e. in order to make is possible to distinguish individual instruments). Owing to the limitations on space in the orchestra pit, which is often overcrowded, it is usually impossible to position the microphones in an ideal stereophonic layout. Generally, however, good results can be obtained by mounting the microphones along the partition between the orchestra and the audience, and dividing them electrically into two groups, the left-hand group being linked via an amplifier to loudspeakers on the left of the stage, the right-hand group via another amplifier to those on the right of the stage. Microphones that have been mounted for this purpose may be seen in fig. 2. Usually the ambiophonic installation will be employed to give additional backing to music originating in the orchestra pit.

Some theatres possess an organ which is sometimes built into a side wall of the stage. A location like this is particularly bad from the acoustical standpoint: most of the sound vanishes into the stage superstructure, which is highly absorbent, and the rest can only reach the auditorium by passing through the wings. It is advisable in fact to "put the organ on stage" in the acoustical sense by employing loudspeakers in support, as is done for the orchestra. Further support, by means of loudspeakers placed around the auditorium and giving ambiophonic (diffuse, delayed and reverberant) sound, makes it possible to render the acoustics of the auditorium almost ideal for organ music [6]).

The acoustics as they affect the performer

What has been said so far concerns auditorium acoustics as perceived by the audience. Acoustics as they affect musicians on the stage or in the orchestra pit are certainly of no less importance. In a good concert-hall each member of the orchestra is enabled by reflections from the wall and ceiling of the podium to hear the playing of the other musicians. In an opera house, on the other hand, it is rare for the stage setting to have acoustically hard walls and a closed ceiling (this would make scene-shifting an extremely heavy job). Opera scenery usually consists of cloth stretched over a framework of laths, and the setting is open at the top, so that acoustically the stage resembles a place in the open air. In consequence, the singers on stage do not hear

[6]) This has been possible in the Philips Theatre at Eindhoven since as long ago as 1954; see fig. 7 in the article cited in [4]).

one another at all well and have no idea of how their combined efforts sound in the auditorium. The musician in the orchestra pit hears little more than the instruments in his immediate vicinity, and he too has no idea of the impression made in the auditorium.

These drawbacks for the performer are considerably alleviated by the electro-acoustical equipment described above; direct sound from the loudspeaker columns beside the stage and indirect sound from the speakers around the auditorium travel directly or are reflected into the orchestra pit and the stage as well as into the auditorium, and in this way the performers get an impression of what is being heard by the audience. The extent to which they do so depends very much on local circumstances. There is in principle a limit to this desirable effect in that the microphones pick up some of the returning sound, with acoustic feedback as a result. Naturally, this must not be allowed to result in accentuation of one or more frequencies. However, one way of obviating the above difficulties for the performer is conceivable, that being to have special loudspeakers for the performers, placed in such a way that the microphones only pick up a very small fraction of the sound from them. It is the orchestra's difficulties that are easiest to remedy in this way, for its members remain in place; small loudspeakers fixed to chairbacks or music-stands, close to the musician's head, serve the purpose quite satisfactorily.

To find a solution for the artistes on the stage who must have plenty of freedom of movement, is not so easy. Here use can be made of loudspeaker columns facing the stage and so positioned that as little as possible of their sound reaches the microphones.

Operation

No extra staff are required for operating the electro-acoustical installations so far mentioned. The microphones and loudspeakers have permanent positions (generally such that they are invisible to the audience, or in any case such that they are inconspicuous). The microphone and loudspeaker cables are likewise permanent (*fig. 3*) and all lead to a control room, which accommodates the pre-amplifiers, power amplifiers and delay device. We shall come back to the control room later on. If reserve cables are laid at the same time as the main installation is being built, it will be easy later on to make extra microphone points in places where these may be desired.

Despite differences in auditorium acoustics and resulting differences between installations, it always proves possible in practice to build up the latter out of standard units, provided these are combined in the right manner. The installation is once and for all adapted to the *auditorium* by the positioning of the loudspeakers, the delays that the ambiophonic equipment has been adjusted to give, and so on. Adaptation to *performances* of various kinds can be obtained by exercising a suitable choice of the intensity ratios and reverberation times; the choice is made by operating a selector switch having positions for "play", "opera", "chamber music", "symphonic music", "soloist concert", "organ", etc. Operating routines can in this way be limited to switching on, carrying out a set inspection programme, and switching the selector to the right position. Such work can be done by a member of the normal technical staff, such as the lighting assistant.

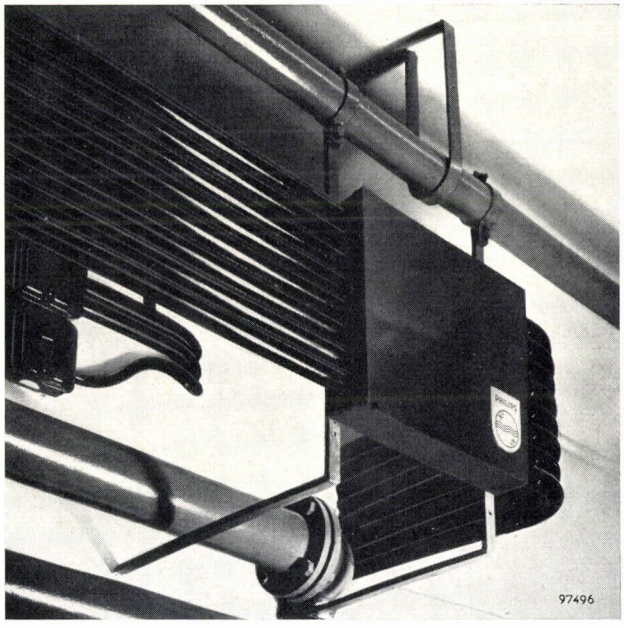


Fig. 3. All the cables of the electro-acoustical installation in the Scala are enclosed in steel conduits.

Electro-acoustical service facilities in theatres

Apart from the above electro-acoustical installations, which have an artistic function, there is a need in theatres for electro-acoustical equipment providing audience, artistes and staff with certain facilities that improve or simplify the everyday routine of the house. Like the others, these "service" installations do not normally require special staff to operate them.

We shall now go on to review a series of installations of this kind.

Monitoring system

If the everyday routine of the theatre is to go smoothly, certain persons outside the auditorium must be able to follow the performance in order to know when to come into action. This applies amongst others to the artistes in their dressing rooms, musicians who have to play behind the scenes during a performance, ballet groups, extras, scene-shifters, ushers, porters, and stage-lighting, kitchen and wardrobe-personnel.

Theatres like the Scala and the Palais de Chaillot have a monitoring system for this purpose. It consists of an amplifier whose input signal is supplied by microphones in the auditorium and which feeds loudspeakers in dressing rooms and other service rooms. The loudspeakers in some of these rooms are fitted with a volume control or on-off switch.

Paging system for performers

One of the responsibilities of the stage manager is to warn actors, extras and musicians a short time before they have to appear or perform. For this purpose use is generally made of the loudspeakers and cables of the monitoring system. Just to one side of the stage, the stage manager has a control desk with microphone and amplifier (*fig. 4*) at his disposal. By means of keys or push-buttons he is able to connect up one or more loudspeakers of the monitoring system to his own amplifier, so that he can address certain individual performers or groups of performers. The employment of a special circuit in which the loudspeaker lines are three-core cables makes it possible for the caller to put the volume controls and on-off switches out of action while he is talking; thus the call is always heard at full volume. The principle of dynamic compression is applied in the amplifier for the paging system; the advantage is that calls made over the system are always equally loud, the volume being independent within wide limits of the distance between caller and microphone, and that reproduction is free from distortion because the compressor makes overloading of the amplifier impossible.

Hearing-aid installation

In a theatre or concert hall, a hearing-aid is not a very satisfactory device for the hard of hearing because it gives no impression of direction. In consequence, the wearer is unable to separate disturbing sound (reverberation and audience noise) from music or the voices of the actors. Hence such noise is far more tiresome for him than for a person with normal hearing. Things are likely to be much improved for the hard of hearing, then, if they can be offered a sound signal that is practically free from interference. This can be achieved by picking up the sounds from stage or platform with directional microphones, so positioned that they are little affected by noise and reverberation from the auditorium, and feeding the signal thus obtained to a loop of wire laid round the auditorium. The audiofrequency current in the loop creates a magnetic field that induces a voltage in the pick-up coils [7]) of normal hearing aids. The voltage induced is to all intents and purposes free from interference, and has the same value throughout the auditorium. With this installation, therefore, there is no need for the hard of hearing to occupy seats in the front of the auditorium, or seats reserved for them and provided with connections for special earpieces. Some theatres having a loop installation hire hearing-aids with a pick-up coil to those of the public who want them and do not possess their own.

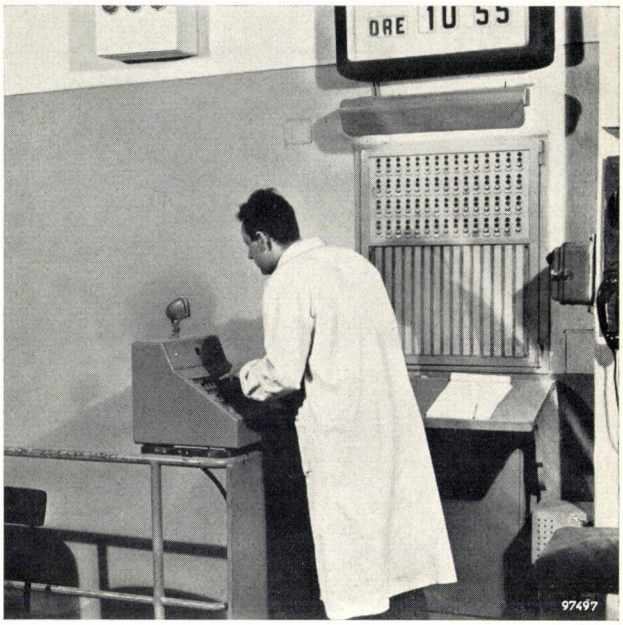


Fig. 4. The stage manager at the Scala has a control desk with a microphone, amplifier and switches; the desk allows him to contact artistes, individually or in groups, via the cables and loudspeakers of the monitoring system, to call them "on stage".

Monitor system for latecomers

It is a well-known fact that, at any performance, some members of the audience arrive late. In large theatres, in order to spare the punctual from the annoyance caused by latecomers, the auditorium doors are closed about one minute before the start of the performance and only re-opened during the short interval that follows the first act or first item on the musical programme. In this way the late-

[7]) Philips tech. Rev. **15**, 41-42, 1953/54.

comers miss a considerable part of the programme. For the benefit of latecomers at the Scala and the Chaillot theatre, the foyers have been equipped with loudspeakers connected to the monitory system described above. During the main interval the loudspeakers can be used for playing back music from gramophone records or magnetic tape. If a foyer is provided in addition with a microphone point and a pre-amplifier, the foyer itself becomes suitable for independent use as a hall where lectures, music recitals and the like can be given.

Television links

Although it is not classifiable as electro-acoustical equipment, closed-circuit television must be given mention here because there does exist a technical connection between the two, and television links can be extraordinarily useful in theatres.

In the first place, television is useful where it is thought desirable to display events on the stage in certain places outside the auditorium. An example is the television installation in the Palais de Chaillot, where latecomers can see the first act of the play from one of the foyers. For this purpose the cinema projection cabin has been equipped with a television camera fitted with a telephoto lens. The cabin also houses a small television transmitter. However, instead of feeding an aerial, the transmitter passes a carrier wave modulated with the video signal into a cable which conveys it to a number of ordinary television receivers in one of the foyers. The installation can, if necessary, be extended to feed extra receivers in other places such as managers' offices, the lighting cabin and so on.

Television has found a different application at the Scala, where it serves to make the conductor visible to members of the orchestra or choir who have no direct view of him. We shall return to this application when discussing that particular theatre.

Paging system for the public

A paging system can be a great help towards keeping a performance running to schedule, particularly in large theatres where during the interval the audience disperses into corridors, stairways and foyers. The system can be used to warn the public that the show is about to begin or that the interval is coming to an end, and to direct latecomers to a particular foyer. If the lines to these loudspeakers pass through the stage manager's control desk, he is able to announce the approach of an interval to latecomers, so that they can make their way to the auditorium. In case of emergency members of the audience can be called by name.

The Palais de Chaillot possesses a particularly convenient system of this kind. Reminders and instructions to the public, interspersed with suitable music, are recorded on tape prior to the performance. The playing back of the tape is initiated a certain time before the start of the show and before the end of the interval, and in this way members of the public are efficiently and affably paged into the auditorium, or into the foyer with the television sets if they have arrived late.

Simultaneous interpreting installations

In virtue of their wide range of facilities, large theatres are highly suitable for congresses. If the congress is an international one it may be essential that translations of speeches and discussions should be supplied in some form or another. Providing oral translations after each speech costs a great deal of time, even if they are nothing but résumés, which in any case have the disadvantage of incompleteness. For this reason the method most commonly used nowadays is to give simultaneous translations in various languages via an installation designed for the purpose.

A row of interpreters' cabins is erected behind a double glass panel looking on to the congress hall (*fig. 5*). By switching a selector, each interpreter

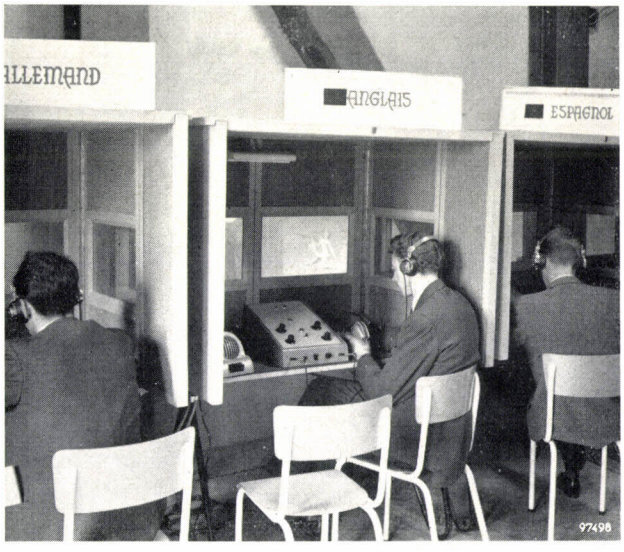


Fig. 5. Interpreters' cabins of the simultaneous interpreting installation in the Grand Auditorium at the 1958 Brussels Exhibition.

can hear in his headphones either the speaker in the hall or one of the other interpreters; he gives his own translation into a microphone. (Sometimes interpreters translate from a language which suits them better than the language of the speaker.)

Two kinds of interpreting installation exist. Under the older system each seat in the hall is provided with a socket for headphones and with a selector for different languages [8]). The need to lay a large amount of cable makes an installation of this kind rather expensive. The "wireless" system is much cheaper; under this system each interpreter's rendering is modulated on a "carrier wave", the set of modulated carriers being passed through one loop laid around the auditorium like the loop in a hearing-aid installation. Each person attending the congress is given a special hearing-aid (in reality a small radio receiver — see *fig. 6*) equipped with a pick-up coil and a selector for the various languages. Thus equipped, participants are able to listen to the original speech (which may be backed up by an electro-acoustical installation) or to the translation thereof which suits them best. The carrier waves can be picked up outside the hall, and the wireless system therefore lacks the advantage of secrecy; under the system first referred to, secrecy is complete. Philips have equipped several buildings with these wireless interpreting installations, some with four channels (the carrier frequencies being 74, 86, 98 and 110 kc/s) and some with six (carrier frequencies of 50 and 62 kc/s plus the four just mentioned). Amplitude modulation is employed. The power output to the loop is generally no more than 2 or 3 W per channel.

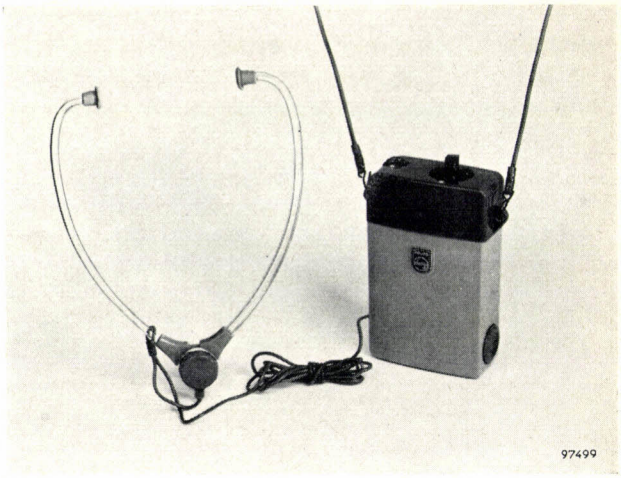


Fig. 6. Personal transistor receiver used in conjunction with a simultaneous-interpreting installation. At the top left is the volume control and top right the channel selector. The installations have either six or four channels.

Equipment in the acoustic control room

Theatre installations serving mainly or incidentally for sound effects call for a special operator who is able to supervise and check the whole electro-acoustical installation from one central point, that being the acoustic control room. The position and layout of the control room should be such as to allow the operator seated behind his control desk to see the stage through a sound-insulating window.

All microphone and loudspeaker lines terminate in the control room (*fig. 7*), where connections to the internal and public telephone systems are available; often there are connections to land lines going to broadcasting and television centres. Besides the central control desk, other desks for special purposes such as radio broadcasts may be present. The control room is equipped with gramophones and magnetic

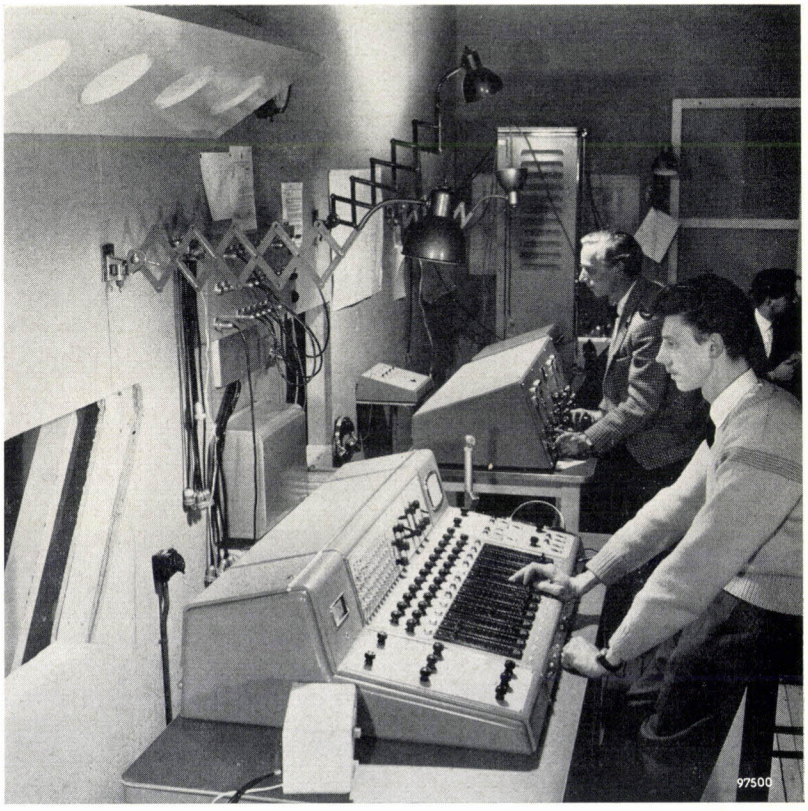


Fig. 7. An example of a sound control room for a large electro-acoustical installation. The photograph shows part of the control room of the Grand Auditorium at the 1958 Brussels Exhibition. The central control desk is in the foreground, the broadcasting desk to the rear. Monitoring loudspeakers may be seen in the top left-hand corner.

[8]) See for example N. A. J. Voorhoeve and J. P. Bourdrez, The electro-acoustic installation in the League of Nations Palace in Geneva, Philips tech. Rev. **3**, 322-330, 1938.

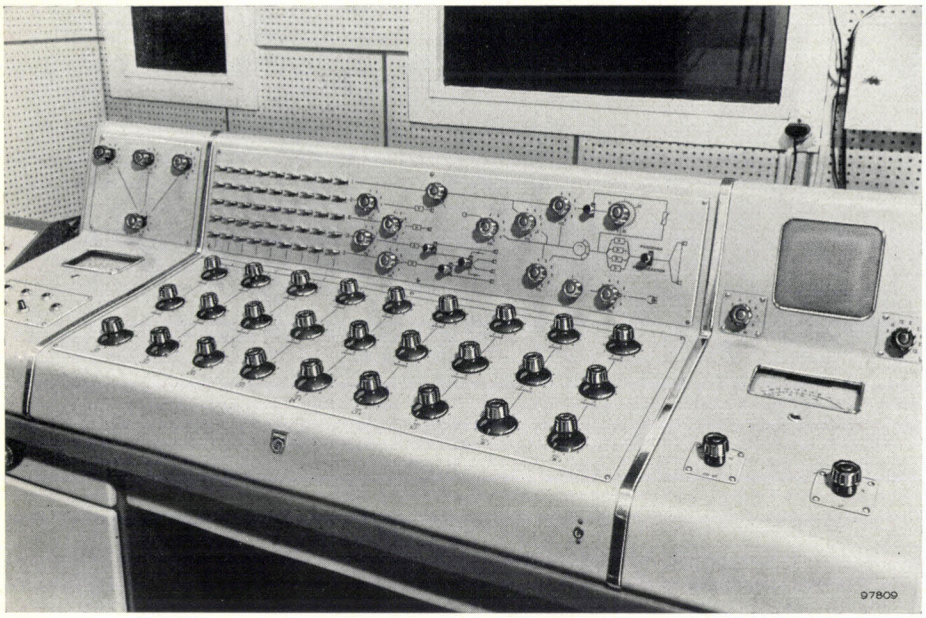


Fig. 8. The central control desk in the Palais de Chaillot. The nearest of the three rows of nine knobs consists of volume controls, the middle and back rows being low-note and high-note tone controls. The sets of three controls, taken from left to right, regulate six microphone channels, one gramophone channel and two tape-recorder channels. On the back panel, at the extreme right of the central section, is the panorama-reverberation switch, discussed later in the article (p. 66).

recording and playback apparatus and it accommodates all the amplifiers, standby amplifiers and monitoring equipment.

Experience has shown that a total of 25 microphone points on and near the stage and in the auditorium is adequate for all eventualities. Some points are in constant use and others are only used temporarily, when certain scenes require them. If the scenes change rapidly, microphones can be plugged in prior to the scene in which they are required; all that is necessary during the change is to place them in position.

It has further been found that the number of microphones in use at any one time rarely exceeds six. However, more than six were necessary on occasion in the Grand Auditorium at the Brussels World Fair (see the description of this installation, below). A switchboard is mounted within reach of the operator, allowing him to switch up to six microphones (or groups of microphones in parallel) through to the control desk. Low-impedance (50 ohm) microphones are employed in theatres, to obviate high-note losses due to the long cables; the microphone inputs in the desk are matched to this impedance.

The gramophones are of the type used in broadcasting studios. They are duplicated in order to allow gramophone records to be played without a break. They have their own amplifiers, which incorporate a correction network designed for the usual record-cutting characteristics [9]). One or other of the gramophones can be connected to an input of the central control desk by means of a selector switch.

The tape-recorders are also of the professional type. Two or three tape speeds are possible and there are generally two channels, in order to allow for stereophonic reproduction. Usually a pair of tape-recorders, together with their amplifiers, are built into a recording desk provided with the necessary record and playback switching arrangements.

The central control desk forms a link between a group of signal sources and a group of amplifier and other inputs. The former group comprises:

the microphones (generally six channels),
the gramophone in operation (usually one channel), and
the tape-recorder used for playing-back (two channels, for stereophony).

The latter group comprises:

the inputs of the power amplifiers driving the various loudspeaker groups,
the input of the ambiophony installation,
the input of the recording desk, and possibly
the input of the broadcasting desk.

On the central control desk (*fig. 8*) may be found all the switches and controls that have to be manipulated during a performance or when a recording is being made. The desk also contains all the pre-amplifiers.

Each input channel of the desk has its own pre-amplifier and volume control and its own tone controls for low and high notes. By switching a selector it is possible to tap any of the channels at a point in front of these controls, and so to listen in to the signal and read its level on a dB meter. The output channels likewise have separate volume controls. Connections between input and output channels are made by means of a cross-bar switchboard, the principle of which is shown in *fig. 9*.

[9]) See for example Philips tech. Rev. **17**, 104, 1955/56 or **18**, 239, 1956/57.

The power amplifiers are accommodated in standardized steel cabinets. Each cabinet holds four amplifiers and one in reserve, and each has a monitoring panel. On the panel are a loudspeaker, an output meter, and a selector switch; with these it is possible to listen in to and measure the output from each amplifier. By pressing a button, the standby amplifier can be put into operation in place of one that developed a fault; the standby is always ready for immediate operation because the heaters of its tubes are kept warm. The changeover is not noticeable in the auditorium.

The ambiophony equipment is also housed in a standard steel cabinet (fig. 23 in Part I). Besides being fed to the loudspeakers around the auditorium, the output from the equipment goes to the controls of the recording and broadcasting desks, so that electrically generated reverberation can be blended into recordings and radio broadcasts.

A small part of the central control desk is generally reserved for the controlling, mixing, measurement and auditory monitoring of the signals originating in the auditorium microphones belonging to the general monitoring system. The output signal from this part of the desk is fed to the amplifiers of the general monitoring system, to the public loudspeakers in the foyer and to the induction loop of the hearing aid installation.

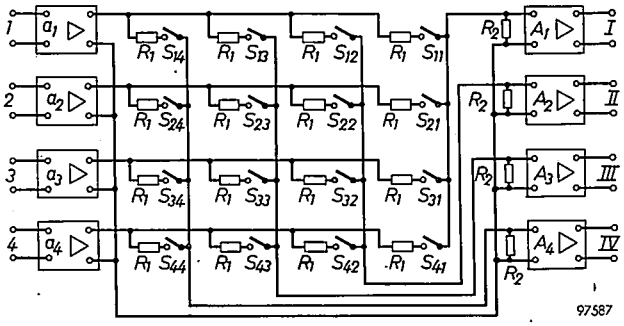


Fig. 9. Block-diagram of a cross-bar switchboard allowing interconnection of multiple inputs with multiple outputs on a central control desk. In the diagram there are four inputs (*1* . . . *4*) and four outputs (*I* . . . *IV*). $a_1 \ldots a_4$: pre-amplifiers. $A_1 \ldots A_4$: line amplifiers. R_1 and R_2 are resistors with high and low values respectively. Suppose (e.g.) that switches S_{23} and S_{43} below are closed: in this case pre-amplifiers a_2 and a_4 will deliver signals to line amplifier A_3 via voltage dividers R_1-R_2. The signals are therefore mixed and amplified on arrival in output channel *III*. The high resistances denoted by R_1 prevent cross-talk between the pre-amplifiers.

Although we have mentioned various theatres by name, we have tried to keep the foregoing account as general as possible. We shall now go on to describe more fully certain large installations already in existence. These descriptions will show that electroacoustical installations, while having the same basic pattern, exhibit individual differences due to local circumstances or personal preferences.

Teatro alla Scala, Milan

The *Teatro alla Scala* was built by Piermarini in 1778 on the former site of a church founded by Beatrice della Scala. The building consists of a large auditorium for operas and concerts, and the "Piccolo Scala", a much smaller auditorium in which chamber music and less elaborate operas are performed and which will not be discussed further here. The large auditorium (fig. 2) has the shape of a horseshoe; it seats about 2500 and there are six tiers of boxes one above the other. The theatre owes its renown as the finest opera house in the world to both the perfection of the performances given there and the magnificence of its stage settings.

In 1954 a committee set up by the management came to the conclusion that a suitable acoustical installation would help to give audiences an even stronger impression of being involved in events on the stage. They further considered that such an installation would also be the answer to the variety of technical problems then existing. The most important of these problems were:

(1) Musical groups and choirs placed at the side of or behind the stage, for the purpose of creating certain sound effects during operatic performances, could not be heard properly in the auditorium because of the damping caused by the scenery.

(2) These musicians and singers were in the way of actors, dance groups and crowds going on and leaving the stage (it is not unusual for more than 200 persons to enter and exit within a short interval!).

(3) Members of the orchestra and choir behind the scenes were unable to see the conductor. The difficulty was overcome by placing them under the baton of a second conductor who had to watch the principal conductor through an opening in the scenery (*fig. 10*).

(4) Actors awaiting their cues behind the scenes were unable to hear the orchestra and follow the stage action properly; consequently it was difficult for them to pick the right instant to go on stage.

It was possible to settle the first three problems once and for all by fitting up a separate studio for musicians and singers not performing on the stage or in the orchestra pit. A closed circuit television link enables these performers to see the conductor. Whenever they are not playing or singing themselves, they can follow the performance via loudspeakers forming part of the monitoring system.

They play or sing when instructed by the conductor to do so and their performance is reproduced through loudspeakers in the auditorium. The lack of space, and the hindrance musicians represented for the actors and extras, and *vice versa*, have been effectively remedied in this way. Moreover, for the purpose of musical performances the studio is acoustically much superior to the confined space behind the stage, where the sound was absorbed by scenery. As the microphones and the loudspeakers

Fig. 10. At one time it was necessary in the Scala to place members of the orchestra behind the stage to provide music "off". Being unable to see the conductor of the orchestra, these musicians had to play under the direction of an "intermediary" who watched the conductor through a hole in the scenery and copied his movements. They now play in a studio where they can see the conductor on a television screen; their performance is reproduced by loudspeakers in the auditorium.

are in different rooms, there is no danger of acoustic feedback; nor, therefore, is there any limitation on the positioning of microphones in the studio and on the sound effects that can be produced there.

Problem number (4), the inability of actors to hear properly while awaiting their cues, has been settled by installing a special monitoring system: microphones trained on the orchestra and the stage have been mounted on the inside of the partition between orchestra and auditorium (fig. 2), and loudspeakers have been set up off stage.

In regard now to the acoustics of the auditorium, it may be noted that the numerous boxes do a great deal to absorb sound. As a result, the reverberation time is fairly short (1.55 s for 500 c/s) when the auditorium is empty and even shorter when it is full. The short reverberation time favours intelligibility, and in this respect the acoustics of the auditorium require no correction. On the other hand there was every reason to employ ambiophony, since it would be desirable not only during symphony concerts but also at times during operatic performances, when a certain scene required it.

In close consultation with the technical staff of the Scala Philips designed an acoustical installation that caters for all the needs of the theatre. Its "première" took place on the 7th December 1955, when Mozart's "Magic Flute" was performed under Herbert von Karajan. All the music formerly played off stage was now relayed from the studio, the musicians there following the conductor via the television link. At the same time large-scale use was made of the facilities for spatial reproduction — stereophony, ambiophony and distant-source effects — that the installation offers.

The installation can be divided into two parts, one for supporting the actual performance and one covering the monitoring system, paging system and other facilities.

Equipment for supporting the performance

Microphones. A total of 25 microphone points have been provided on the stage and in the studio. Sensitive electrodynamic microphones of type EL 6040 are plugged into these points as required.

Acoustic control room. The complete block diagram of the acoustical installation in the Scala appears in *fig. 11*. *Fig. 12* shows what combinations are possible between the various groups of microphones and loudspeakers. Most of the equipment shown in these diagrams is accommodated in the control room, located between the small and large auditoriums and with view of both stages.

The 25 microphone points on the large stage lead to a switchboard D (fig. 11). Following the instructions in the sound scenario, the operator uses this panel to connect up the microphones M_1 about to be used, with the microphone inputs (which are five in number) of the central control desk. The latter also has a tape-recorder input T and a gramophone input G. The main uses to which the tape-recorder is put are to play back sound effects and to record ballet music that is required for later reproduction at rehearsals or for instructional purposes in the ballet school. The tape-recorder forming part of this installation has only subordinate importance and is not therefore designed for stereophony.

Each of the seven channels M_1, T and G has two pre-amplifiers between which there is a volume control P_1. These controls are not operated from the control room, but are mounted on a correction desk under the control of the sound *régisseur* (see below).

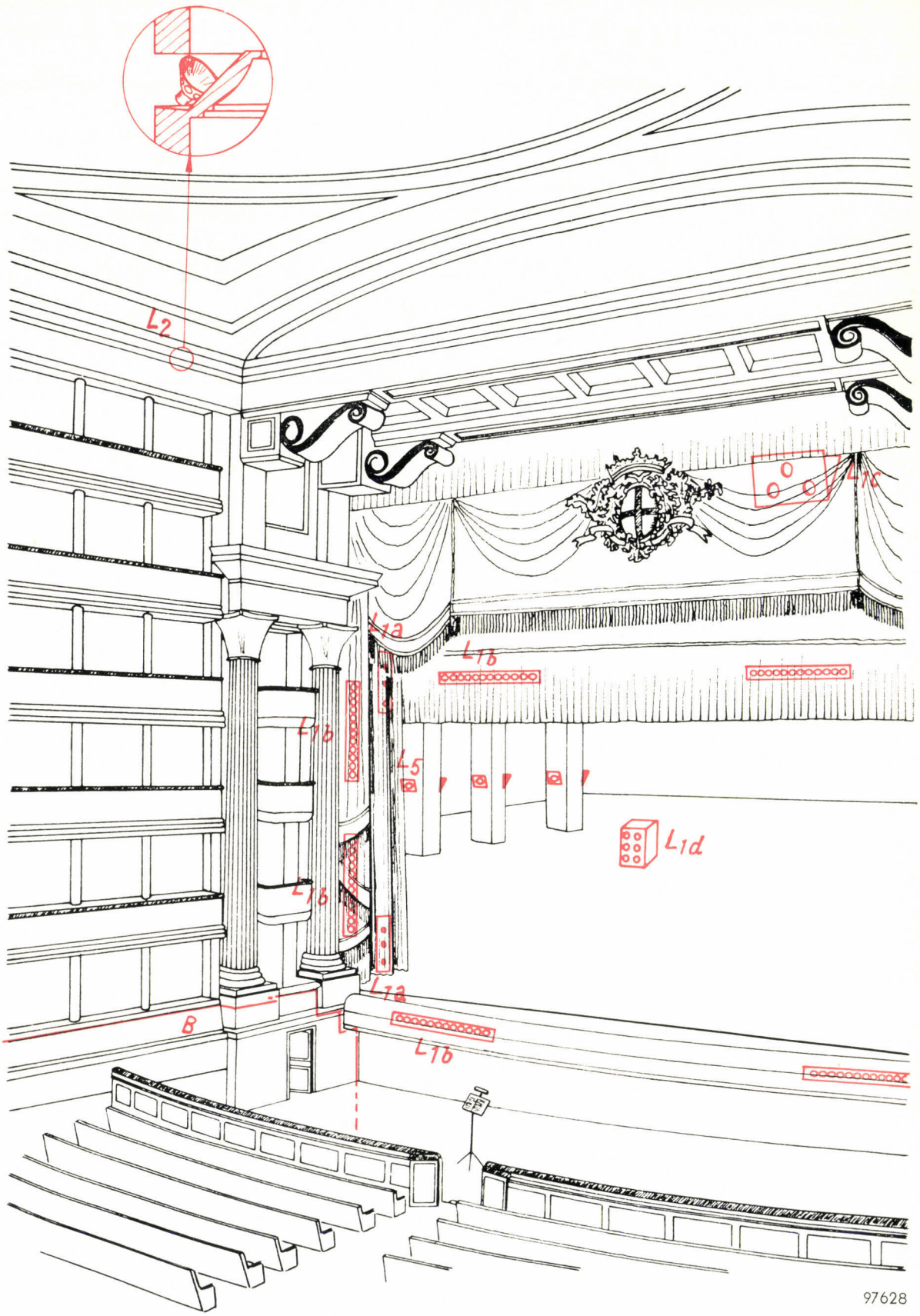


Fig. 13. Location of loudspeakers in the large auditorium of the Scala. L_{1a} are low-note boxes and L_{1b} loudspeaker columns mounted around the proscenium opening and providing panoramic sound. Placed above the stage, loudspeakers L_{1c} are used for overhead sound effects. Loudspeaker unit L_{1d} can be moved about the stage. Loudspeakers L_2 are mounted in the cornice and are responsible for ambiophony. L_5: loudspeakers for the benefit of actors awaiting their cues in the wings. B: induction loop of hearing-aid installation.

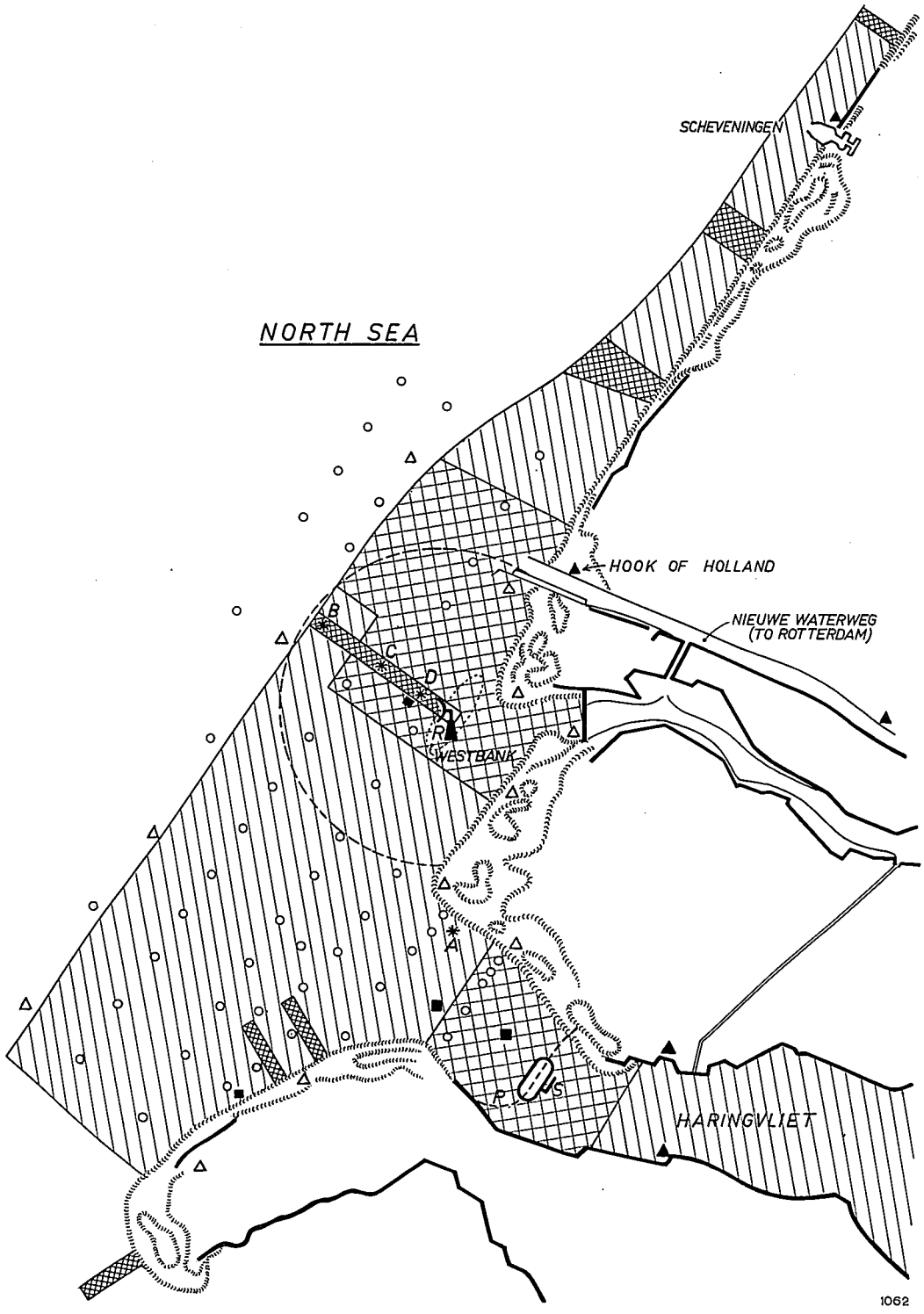


Fig. 2. Survey of the programme of measurements to be carried out during and after the damming of the Haringvliet. In the hatched sea-areas soundings will be made at regular intervals: ▨ once a year, ▩ several times a year, ▩ more frequently still. The symbols ▲ and ■ denote permanent recording stations for the tides, and for waves and tides, respectively. o places where currents will be measured; ✳ *ABCD* tracer dumping sites around which sand movements will be tracked by activity measurements; △ temporary stations for tidal measurements; *R* radar station; *S* sunk foundation for sluices in the Haringvliet; *P* course of dam proper.

The control room also houses an ambiophony installation R (with microphones M_2), 18 power amplifiers (not including standby amplifiers) each of which has a power output of 120 W, and a number of monitoring loudspeakers.

The control room is linked to the sound *régisseur* and the studio by an intercommunications system. Musicians in the studio are able to follow the stage performance via loudspeakers connected to the monitoring system. The loudspeakers in the studio are put out of circuit whenever the microphones are switched on. Red pilot lamps in the studio and the control room and near the conductor and sound *régisseur* indicate that the microphones in the studio have been switched on.

Correction desk. The lighting and sound *régisseur* is responsible for lighting and electro-acoustic matters during the performance. His post is behind the Presidential box, where he sits at the correction desk referred to above; the desk enables him to correct the volume levels to which input channels M_1, T and G (fig. 11) have been adjusted, and to switch off the various groups of loudspeakers by means of relays.

Loudspeakers in the auditorium and on stage. The positioning of the various loudspeakers in the auditorium and on stage may be seen from *fig. 13*. A horizontal and a vertical loudspeaker column L_{1b} are placed in each of the four corners of the proscenium arch; associated with each pair of columns is a loudspeaker box L_{1a} for low notes. Individual loudspeakers belonging to the columns L_{1b} are aimed in directions such as will result in the high notes being distributed as evenly as possible. In combination, the loudspeakers of each pair of columns cover every seat in the house. The high-note loudspeakers (i.e. the columns) and the low-note loudspeakers are driven from separate power amplifiers. (This is the "Bi-Ampli" system [10]). The amplifiers are connected to separate output channels of the central control desk (fig. 11). The desk is provided with a control P_3 for adjusting the overall volume from all four loudspeaker groups. There are two further controls, P_4 and P_5, which, without altering the overall volume, govern the balance between the upper and lower groups and that between the left-hand and right-hand groups. It is possible by means of controls P_4 and P_5 to shift the stereophonic sound image vertically and horizontally through the proscenium arch.

[10]) The advantages of reproducing low and high notes via separate channels have been described by 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, and in particular p. 304 and p. 311 et seq.

Loudspeakers L_2 are fixed in a cornice running around the ceiling of the auditorium (fig. 13). These give diffuse sound, and can be used in two different ways. When they are connected up as a group the sound from them, though diffuse, clearly comes from above; this may be useful for certain effects. Connected to the four channels of the ambiophony installation, the loudspeakers can be made to simulate desired acoustic properties and to give the impression of a hall of any desired size; further, if their volume is given the right relationship to that from loudspeakers L_{1a} and L_{1b}, or the loudspeakers L_{1d} located on the stage, a certain impression of remoteness can be created. In either case a special effect can be obtained by using only those loudspeakers of the L_2 group that are situated at the back of the auditorium. It is possible thus to suggest sound echoing back from mountains, for example, or to bring about an interplay of sound with the stage (panoramic sound).

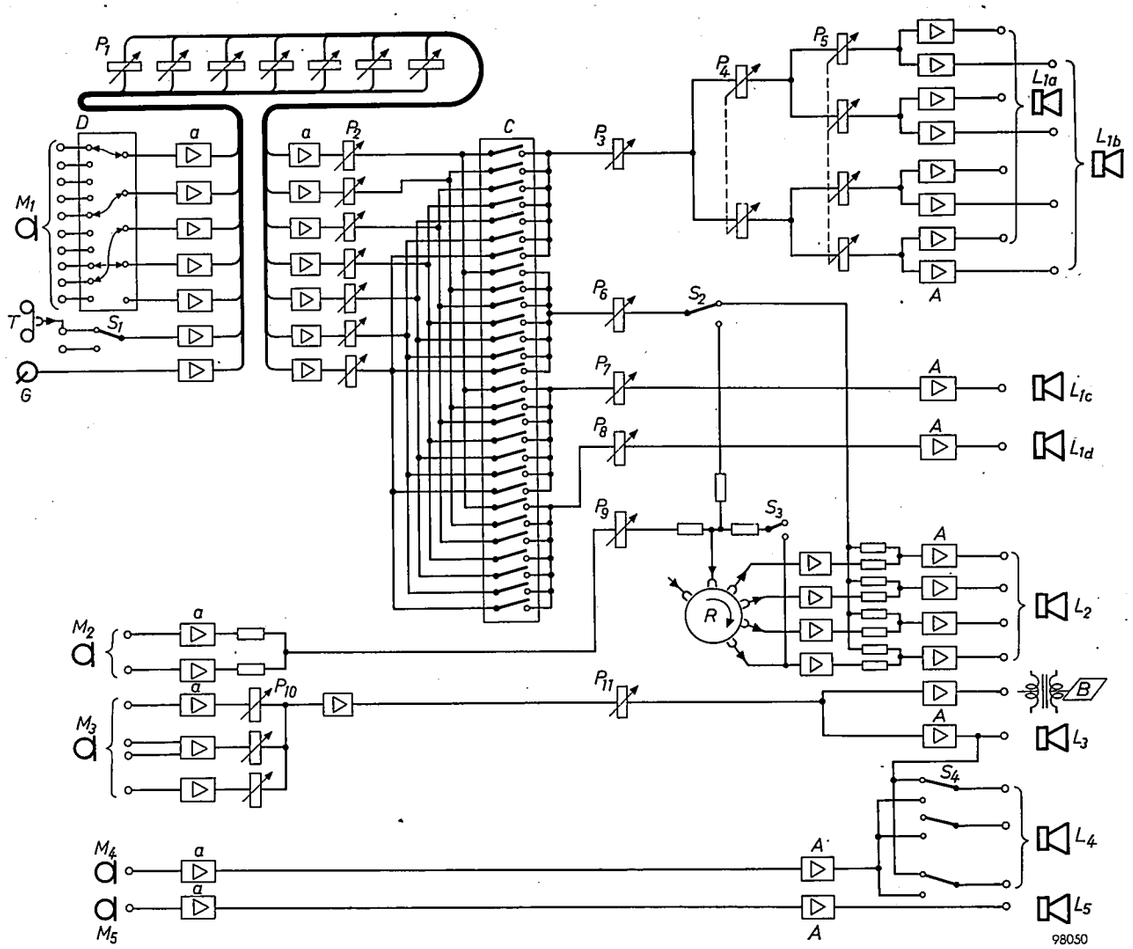
A control P_9 on the central control desk (fig. 11) enables reverberation to be added to the sound which is coming form the stage and being picked up by microphones M_2 above it. Control P_6 adds reverberation to sound coming from the studio or from recordings.

The loudspeakers L_{1d} just referred to as being on the stage (fig. 13) are mounted in groups on trolleys, each group having an output of 60 W. Their purpose is to reproduce instrumental and choral music and sound effects and the like from some point in the wings. Finally, L_{1c} is a loudspeaker group above the stage, whose function is to reproduce overhead sound effects — storms, rain, choirs of angels and the like. The volume controls for L_{1c} and L_{1d}, P_7 and P_8 respectively (fig. 11), are on the central control desk.

All the loudspeakers in the auditorium are of the double-cone EL 7021 type. The group around the ceiling and each of the four groups inside the proscenium arch are capable of producing, without perceptible distortion, a loudness level of 94 phons in the auditorium. This means that the music of a large orchestra or choir can be reproduced at the natural loudness level.

Monitoring and paging system; induction loop

The signal for the monitoring system is derived from four microphones M_3 (fig. 11) which, together, give an overall impression of the performance. One is placed behind the hangings above the proscenium, two in the orchestra pit and one in the chandelier high up in the auditorium. The signals from these microphones are mixed on a separate panel on hte



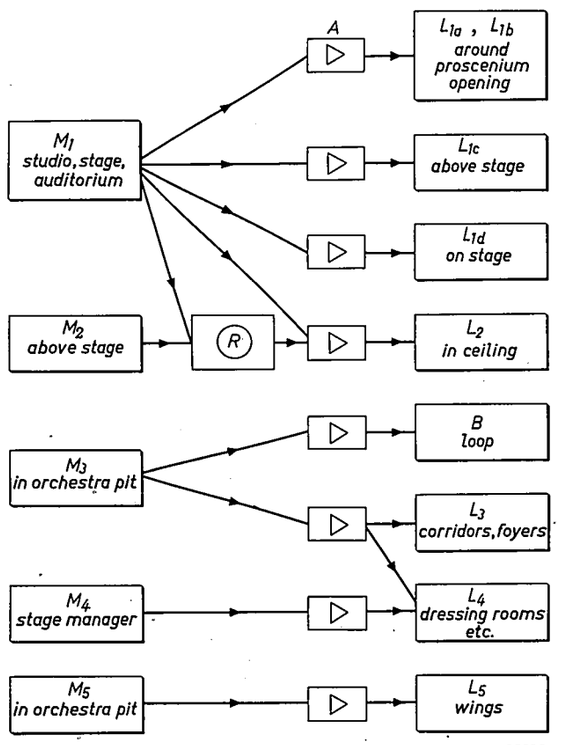


Fig. 11. Block circuit covering the entire electro-acoustical installation (except the intercommunications system) in the Scala.
a: pre-amplifiers. *A*: power amplifiers (120 W). *B*: induction loop for the hard of hearing. *C*: cross-bar switchboard. *D*: microphone switchboard. *G*: gramophone pick-up. L_{1a} and L_{1b}: low-note and high-note loudspeakers mounted around the proscenium arch. L_{1c}: loudspeakers mounted above the stage. L_{1d}: loudspeakers mounted on the stage. L_2, loudspeakers round the ceiling. L_3: loudspeakers in corridors: foyers etc. L_4: loudspeakers in dressing rooms etc. L_5: loudspeakers in the wings. M_1: microphone points on stage and in the studio (only 10 are shown out of a total of 25). M_2: microphones above the stage, for reverberant sound. M_3: monitoring microphones. M_4: stage manager's microphone (fig. 4). M_5: microphones in the orchestra pit (fig. 2). P_1: volume controls on the sound *régisseur's* correction desk. P_2: volume controls in the input channels of the central control desk. P_3: volume controls for loudspeakers L_{1a} and L_{1b}. P_4,P_5: panoramic potentiometers (L_{1a},L_{1b}). P_6: volume control for mixing reverberation into recordings or into the signal from M_1. P_7: volume control for loudspeakers L_{1c}. P_8: volume control for loudspeakers L_{1d}. P_9: ambiophony volume control, associated with M_2 and L_2. P_{10},P_{11}: volume controls for monitoring system and induction loop. R: delay wheel. S_1: selector for a second tape-recorder. S_2: direct/delayed sound switch for L_2. S_3 switch in feedback path of delay equipment. S_4: keys on stage manager's desk. T: tape-recorder.

Fig. 12. Diagram to show how microphones, delay equipment, amplifiers and loudspeakers can be combined in the Scala. The letters have the same meanings as in fig. 11.

central control desk, and are reproduced through the loudspeakers of groups L_3 and L_4 in the sound control room, in the 60 dressing rooms, in the foyers (for the benefit of latecomers), in the offices of the management and in the studio.

Behind the stage the stage manager has his own control desk (fig. 4) with a microphone M_4 and amplifier (figs. 11 and 12). By manipulating the keys S_4 of this desk he is able to call up individual artistes or groups of artistes via the loudspeakers of the monitoring circuit.

For the convenience of artistes awaiting their cues in the wings, loudspeakers L_5 have been suspended from the pillars on either side of the stage (fig. 13). The signal for these is derived from two microphones M_5 (fig. 11) in the orchestra pit; it passes through an amplifier with a compressor circuit, which obviates overloading during fortissimo passages, though solos and pianissimi remain quite audible.

An induction loop is laid round the auditorium (B in figs. 11 and 13) for the benefit of the hard of hearing. The signal is derived from M_3, the microphones feeding the monitoring circuit, and passes through a 120 W amplifier. Hearing-aids are hired out by the theatre.

Théâtre National Populaire (Palais de Chaillot), Paris

The Palais de Chaillot was built for the World Fair held at Paris in 1937, and occupies the site of the old Trocadero. It consists of a combined theatre and concert hall seating 2900, part of which is built into rising ground, and of two wings which lie at a higher level and which are used as museums. In front of the stage is a spacious orchestra pit. The rear wall of the stage is entirely occupied by a large organ. The theatre is used for stage performances (mainly given by the Théâtre National Populaire company under Jean Vilar), symphony concerts, organ recitals, film shows and galas (shows of various kinds to which admission is by invitation only); on more than one occasion the theatre has served as a congress hall for the United Nations.

The reverberation time is on the short side for music, but for speech it is so long that intelligibility suffers. Many of the seats are so far from the stage that intelligibility is definitely poor.

The existing electro-acoustical installation was built on the initiative of the technical management of the Théâtre National Populaire, who were conscious of the need to settle the intelligibility problem once and for all, numerous experiments having yielded no solution. The decision in regard to the equipment to be used was certainly not unconnected with the success of the "Spectacles Son et Lumière", for which numerous French châteaux and other historical buildings had been provided with electro-acoustical installations.

In the Théâtre National Populaire, unlike the Scala, the emphasis lies on improving intelligibility. The preference for recorded music reproduced stereophonically on the stage, instead of live music relayed from a studio, constitutes a second difference. The arrangements for recording sound are accordingly more elaborate than those at the Scala. Provision has also been made for improving the acoustics of the theatre when symphonic concerts and organ recitals are given. Something has already been said about the sound and television links to the foyer, intended for latecomers, as also about the paging system used to announce the end of the interval and the like, to members of the public gathered in foyers and corridors and on staircases etc. In addition to this, there is an monitoring system for actors in their dressing rooms and for members of the technical staff, the latter being in touch via an intercommunications system.

Despite the above differences in the emphasis laid on various aspects of the installation, in the technical sense it bears a strong resemblance to that in the Scala. The heart of the Chaillot installation is the acoustic control room housing the central control desk (fig. 8). But for a few points of detail, the control room is similar to that in the Scala. One difference is that the Chaillot installation allows for stereophonic reproduction of direct sound; there are two quite separate channels for the loudspeakers beside the stage, and two further separate channels for the loudspeakers on the stage. Moreover, the loudspeakers around the auditorium can be used to create moving and panoramic sound images. The presence in the Palais de Chaillot of a special recording desk is evidence of the prominent place given to recorded sound. The desk incorporates two tape-recorders for recording and playing back, and two gramophones, all of professional type. Because two tape-recorders are available, it is possible not only to give a continuous programme of recordings, but also to make copies on the spot, and to produce special effects by transposing speeds (as is done by composers of electronic music [11]) and in other ways.

Microphone positioning

25 microphone points are provided for the stage. The sound *régisseur* is free to decide on the layout of

[11]) H. Badings and J. W. de Bruyn, Electronic music, Philips tech. Rev. **19**, 191-201, 1957/58.

microphones in the orchestra pit, for the organ, and on the stage during galas. There is little danger of acoustic feedback from microphones employed thus, but stage performances are another matter. Microphone positioning then becomes much more difficult. The microphones must be placed unobtrusively, they must not interfere with stage action and scene-changing, nor must they pick up any extraneous noise from behind the scenes. In many respects the best place is in the footlights, where the microphones can catch high frequencies in the speech of actors facing the auditorium, as is needful for proper intelligibility. The difficulty is that the microphones have to pick up sound from the whole of the stage (which is 50 feet deep and up to 75 feet wide!); at the same time the position in the footlights is an extremely bad one from the viewpoint of acoustic feedback, for the microphones are right in front of the loudspeakers which stand on either side of the stage and which have to render the sound intelligible at the back of the 40 m long auditorium. Clearly, the problem of acoustic feedback dominates everything else here. Indeed, the best possible arrangement of the microphones was only discovered after a great deal of trial and error. Now the microphones are always set up in the same places and their signals are always combined in the same way.

Four small columns, each containing four electrodynamic type EL 6040 microphones, are placed in the footlights. The columns are in symmetrical pairs to the left and right of the centre of the stage. The left-hand pair feeds a loudspeaker column on the left of the stage, the right-hand pair a loudspeaker column on the right of the stage, this providing a stereophonic effect. With this installation intelligibility is good throughout the auditorium, there are no unwanted echoes, and the audience are not conscious of the fact that any of the sound is coming from the loudspeakers.

Stereophonic recordings

Stereophonic recordings of music on the stage are made with the aid of an artificial head. There is usually no audience on these occasions, and the appearance of the head would not have mattered were it not for the fact that it is occasionally desired to make recordings (of a documentary nature, for example) in the presence of the public. The artificial head had therefore to be as inconspicuous as possible, and accordingly it was made in the form of a transparent plate with a slim microphone of type EL 6040 on either side. Part of the plate can pivot in its own plane, allowing large or small differences to be introduced into the path lengths to the two microphones; these differences correspond to stereophonically-effective field angles [12]) of 90° and 180°. The pivoting portion is adjusted to the angle that gives a stereophonic sound image corresponding as closely as possible to the configuration of the various musical instruments during the recording. For recording a large orchestra or a stage performance, the artificial head is suspended in front of the stage; for recording stereophonic sound effects or the music of a small ensemble, the head is mounted on a stand and placed on the stage.

The auditorium and stage loudspeakers

The total of 129 loudspeakers employed for improving the acoustics of the theatre can be divided into three groups according to location, namely (*a*) around the proscenium arch, (*b*) on the stage, and (*c*) around the ceiling.

(*a*) *The loudspeakers around the proscenium arch* are in two symmetrical groups to left and right of the centre of the stage. Each group comprises a vertical column beside the proscenium arch, a horizontal column placed at the top (see figs. 11 and 12 of Part I) and a box for low notes. The left-hand and right-hand groups are connected to separate output channels of the central control desk.

Each of the *vertical columns* comprises 18 double-cone loudspeakers of type EL 7021, which reproduce frequencies above 250 c/s. They are inconspicuously housed in small cabinets aimed towards the audience, so as to provide direct sound. Successive loudspeakers forming a column are mounted at slightly different angles so that, in combination, they cover the whole of the auditorium. This results in even diffusion of the high notes (see Part I, p. 319) and makes the radiating angle of the columns almost independent of frequency.

Each of the *horizontal columns* comprises ten loudspeakers of type EL 7021, housed similarly in small cabinets. The angles at which they are mounted are also staggered, but they are pointed towards the ceiling and the walls, not towards the audience. In this way maximum diffusion is achieved.

Each of the *low-note boxes* contains two EL 7031 high efficiency (14%) loudspeakers that reproduce frequencies below 250 c/s. The cubic capacity of the boxes is large enough to allow them to reproduce the lowest musical notes of all. They are mounted near the ceiling, above the vertical columns.

The central control desk enables the above groups of loudspeakers to be operated in the following five ways:

---

[12]) Philips tech. Rev. 17, 174 (fig. 5), 1955/56.

(1) the vertical columns only, for speech;
(2) the vertical columns combined with the low-note boxes, for music (direct sound);
(3) the horizontal columns only, for diffuse sound from above the stage;
(4) the horizontal columns combined with the low-note boxes, for music (diffuse sound);
(5) all groups together, for the optimum reproduction of music.

The optimum relationship between the sound volumes given by the vertical and horizontal columns and the low-note boxes was established from comprehensive listening tests.

During stage performances at the Théâtre National Populaire the spoken word is supported stereophonically via the microphones in the footlights and the vertical loudspeaker columns. Stereophonic recordings of music and sound effects are reproduced by one of the five loudspeaker combinations listed above.

(*b*) *The loudspeakers on the stage* are housed in large cabinets; each cabinet contains one EL 7031 and four EL 7021 loudspeakers, the combined output being 60 W. Their function is to reproduce music and sound effects from the wings during stage performances. They can be connected to two separate output channels of the central control desk, and it is therefore possible to arrange for two different sounds to come simultaneously from different directions, or to achieve stereophonic reproduction either in breadth or in depth. Combined with the horizontal columns, the stage loudspeakers can be made to create a panoramic effect whereby the sound appears to travel around the proscenium arch; by combining them with the vertical columns it is possible to move the sound image up and down stage.

(*c*) *The loudspeakers around the ceiling* are of type EL 7021 and are housed in shallow cabinets aimed obliquely upwards, so that they give diffuse sound. Diffuse reverberation can be produced by conferring suitable delays on the signals to these loudspeakers; an EL 6910 ambiophony installation embodying a magnetic delay wheel (see figs. 22 and 23 in Part I) is employed for this purpose. In principle, the ceiling loudspeakers are apportioned amongst the four delayed-signal channels on a random basis, but care is nevertheless taken to see that no speaker is given a delay less than that consonant with its distance from the stage when the delay wheel has a peripheral speed of 3 m/s. If the speed is reduced to 1.5 m/s all the delay times are doubled, and the auditory impression is that of a hall of very great size. The central control desk has facilities for giving reverberation times between 0 and 4 sec to the sound from the ceiling loudspeakers.

During concerts the signal for the ceiling loudspeakers is derived from microphones above the orchestra or at the organ; during stage shows and galas the ceiling loudspeakers are supplied from microphones in the footlights or on the stage, or from tape-recordings or gramophone records. These loudspeakers can be combined with the stereophonic groups, or with the large units on the stage, in order to produce distant-sound effects.

The earlier-mentioned random distribution of the ceiling loudspeakers amongst the four delay channels — necessary for ambiophony — is obtained only when a certain switch on the central control desk is in the "reverberation" position. When the switch is put into the other position, marked "panorama", relays come into action and modify this random distribution, connecting each of the four amplifier outputs to a group of loudspeakers occupying one quadrant of the ceiling. At the same time other relays bring a "panorama potentiometer" into circuit; this device, which is also housed in the central control desk, now controls the inputs of the four amplifiers. The signals to the ceiling loudspeakers are no longer being delayed, and consequently the sound from these speakers gives an impression of direction that depends on the proportions in which the four quadrant groups are contributing to the overall volume. The panoramic potentiometer enables these proportions to be varied, and in consequence the apparent direction of the sound to be altered. Very striking sound effects can be produced in this way. The signal for panoramic sound is of course derived from a recording.

As an illustration of the scale of the "service" installations (monitoring, paging and intercommunications), it may be noted that they involve the use of 355 loudspeakers outside the auditorium.

Grand Auditorium at the 1958 Brussels World Fair

For the purposes of the international exhibition held in Brussels in 1958, one of the "palaces" that had survived from the 1935 World Fair was converted into a large hall (fig. 1 — in the meantime it has been demolished). The hall seated 2300 persons and was used for the many shows and gatherings already mentioned (p. 53).

The hall had a cubic capacity of 15 000 m³, its walls were covered with a material that was highly sound-absorbent, and the lighting arrangements required that it should have a perforated ceiling. All this resulted in a reverberation time (1.4 sec at 500 c/s in the empty hall) that was far too short for music. The acoustics were not unfavourable to

speech, but intelligibility left much to be desired at the back of the hall, owing to its great length (45 m).

Intelligibility

Intelligibility was improved by means of the stereophonic support afforded by 2×2 microphones on the edge of the stage and loudspeaker columns on either side of the proscenium arch (fig. 1). It was impossible for architectural reasons to give individual loudspeakers directions such that the high frequencies would be evenly distributed throughout the hall. Moreover, it was found that, in some parts of the hall, the position of the stereophonic image varied according to the seat occupied. These difficulties were overcome by placing small loudspeaker columns behind the hangings above the stage, giving them suitable angles and connecting them in parallel with the vertical columns.

As was shown in Part I (p. 319), it is desirable to attenuate the low frequencies in speech; this was achieved with the aid of a high-pass filter with a 250 c/s cut-off.

Music from the orchestra pit

The same stereophonic installation, but with 2×2 microphones placed in the orchestra pit, served for supporting the music played in the pit, insofar at least as frequencies above 250 c/s were concerned. For reproducing the range below 250 c/s, low-note loudspeakers, placed beside each of the vertical columns, were switched in via the central control desk; further there was a choice between low and high positioned groups for the low notes. The low-note loudspeakers were driven via a low-pass filter with a 250 c/s cut-off.

This installation was also used for backing up singing on the stage and the quieter instruments played there.

Acoustics for music

The points and circles plotted in *fig. 14a* indicate values of the reverberation time T measured by the Technical-Physics Service of the T.N.O. (Applied Physics Research Institute) and the Technische Hogeschool Delft with the ambiophony installation out of action; see also curve *1* in fig. 14*b*. The reverberation time T was 1.4 sec at 500 c/s and 0.6 sec at 10 000 c/s. In a hall with a cubic capacity of 15 000 m³, these values should be about 2 sec and 1.0 sec if the acoustics for music are to be good. After final adjustments of the ambiophony installation and application of a correction for the low-notes, T varies approximately according to curve *2* with the reverberation knob turned up to the maximum extent, and according to curve *3* with the knob in a central position. The knob was placed in a central position for most of the musical programmes, with very satisfactory results.

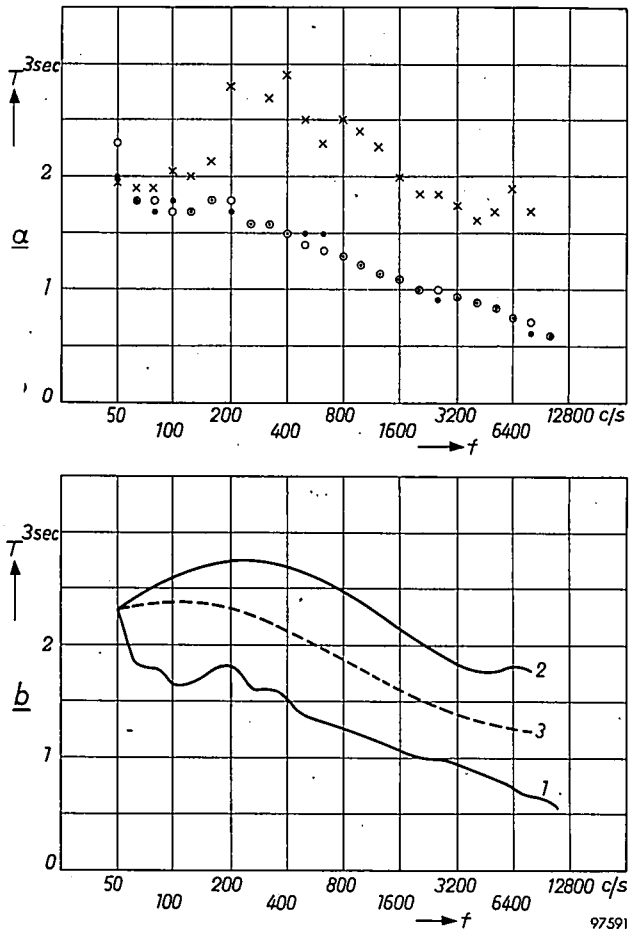


Fig. 14. Reverberation time T as a function of frequency. The diagrams relate to the Grand Auditorium at the 1958 Brussels Exhibition, with no audience present and the curtains drawn. (*a*) Values measured by members of the Technical-Physics Service T.N.O.-T.H., Delft. The points and circles represent values measured with a microphone in the front and rear of the auditorium respectively, and with the ambiophony installation out of action. The crosses relate to provisional measurements carried out with the ambiophony installation giving maximum reverberation, but before final adjustments had been made to it.
(*b*) Curve *1* is drawn through the points and circles in *a*. Curves *2* and *3* were derived by the author after the ambiophony installation had been finally adjusted and correction had been made for low frequencies. Curve *2* represents T as a function of frequency f when the reverberation control was turned up to the maximum, and curve *3* is appropriate to the central position in which the control was placed for most of the musical performances.

Broadcasts

A very large number of the events taking place in the Grand Auditorium were broadcast. The existing microphones of the installation were employed for this purpose. All these microphones were of the condenser type, that type having been chosen with

the requirements of radio broadcasting in mind. The condenser microphone is remarkable for having a very flat response curve that extends from the lowest to the highest sound frequencies. The directivity pattern of EL 6051, the type with which the hall was equipped, can be changed by switching; the directional characteristic can be either omnidirectional or cardioid. (cf. Part I, p. 326).

The positioning of microphones for improving the acoustics of the hall was based as far as possible on the layouts required by the broadcasting engineers. Signals for the broadcasting desk were derived from the six microphone channels on the central control desk (tapped from points *before* the controls, to permit separate adjustment of the sound in the hall and the broadcast signal).

In some cases layouts other than the broadcasting ones were considered desirable from the viewpoint of auditorium acoustics. It happened in this way that on occasion more than six microphones were needed at a time. This situation called for an extra control desk allowing six microphones, whose signals could be adjusted separately, to be connected up to *one* input of the broadcasting desk. Thus the broadcasting engineers had a total of eleven microphone channels at their disposal. Electrically generated reverberation was available for mixing into the radio signal.

Radio commentators had the use of four booths with a view of the hall.

Other facilities

The other facilities can be dealt with very briefly. Much importance was attached to the recording of performances and the reproduction of recordings on tape or discs. Accordingly, special desks were available for gramophone reproduction and for stereophonic recording and reproduction using magnetic tape, the former desk embodying two professional gramophones and the latter two professional tape-recorders. It was possible to blend electrically-generated reverberation into the recordings, the central control desk being employed as a mixing desk.

Most performances in the Grand Auditorium were given once only. Rehearsals were therefore the exception rather than the rule, and they were always limited, so there could be no question of carefully preparing the sound effects. Accordingly, no arrangements were made for providing panoramic sound, as available in the Scala and the Chaillot theatres. Even so, certain sound effects over and above stereophony and ambiophony were available with the installation as it stood. By cutting out the filter for the low-note speakers referred to above, and using each of them separately to reproduce the whole audible range, it was possible to cause sound to come from left or right or from above or below: operated in conjunction, these speakers could be made to suggest a sound source moving horizontally or vertically.

The "wireless" installation for simultaneous interpreting, with its six interpreter's cabins and six channels, has already been discussed (figs. 5 and 6). Mention may also be made of the loop installation for the hard of hearing and of the general monitoring system for artistes in their dressing rooms, technical personnel and members of the public in foyers.

Gebouw voor Kunsten en Wetenschappen in The Hague

Dating from 1874, the Gebouw voor Kunsten en Wetenschappen (Arts and Sciences Building) contains a hall seating 2100 and having three balconies. The hall is mainly used for concerts, but in addition plays and shows are given and big meetings are held there. It has possessed an ambiophony installation since 1954. Apart from the Philips theatre in Eindhoven, this was the first building to be provided with such an installation [13]). Experience gained here has contributed a great deal to the success of the installations built in other countries and described above [14]).

In October 1958 the old installation made way for a new one. The reasons for the change were (1) it was desirable to replace the prototype ambiophony installation by the definitive version of the equipment; (2) the large loudspeakers mounted along the parapet of one balcony were unsightly; and (3) further steps had to be taken to improve intelligibility and to "bring out" the organ, which was badly sited in the acoustical sense.

Here we shall do no more than touch upon a few particular aspects of the new installation.

Loudspeakers for ambiophony

The old ambiophony installation comprised 54 loudspeakers mounted in the rear of the hall and in the back walls of the first and second balconies, and six large boxes fixed to the parapet of the second balcony. The latter were so bulky and conspicuous that it was felt that they should not figure in the final

[13]) See figs. 3, 5 and 8 and p. 265 of the article cited in [4]).

[14]) We should like to take this opportunity of expressing our appreciation for the interest and cooperative spirit that the management of the Gebouw voor Kunsten en Wetenschappen have always shown in connection with experiments in this new field of acoustical engineering.

installation. Necessarily bulky, because good low-note reproduction requires a large enclosed volume of air behind the cone, these loudspeakers could not be rendered inconspicuous except by mounting them high up in the side walls of the hall, immediately under the ceiling. But this positioning would have defeated the main aim in view, in that the high notes would not have penetrated sufficiently far into the hall and there would have been places where the sound was not diffuse enough.

We therefore adopted a quite different approach, based on the fact pointed out in Part I that it is impossible to tell the direction from which low-pitched sounds are coming. The following solution was settled on. To augment the sound from the 54 loudspeakers referred to, which reproduce both high and low frequencies, eight large loudspeakers handling low frequencies only were mounted in the side walls just under the ceiling. High frequencies are reproduced by a total of 106 loudspeakers in small cabinets scattered over the walls. 52 of these are inconspicuously mounted beneath the first balcony and along the lower edges of that and the second balcony (*fig. 15*). The cones of the loudspeakers along the balconies are inclined downwards, giving coverage to the stalls. In spite of being angled on to the arena of the hall, these loudspeakers produce diffuse sound in virtue of their large number and their random distribution amongst the four delayed-signal channels of the EL 6910 ambiophony installation. The cabinets are open at the back, so that sound radiated to the rear contributes to reverberation on and below the first balcony. The four delayed-signal channels drive the low-note and high-note loudspeakers via filters with a cross-over frequency of 250 c/s.

16 high-note loudspeakers in small cabinets are mounted underneath the first balcony, 16 on that balcony, 12 on the second balcony and 10 on the third one. Besides forming part of the ambiophony arrangements, these loudspeakers have an important function in the installation for improving intelligibility, which we shall now go on to discuss.

Improving intelligibility

Earlier on, intelligibility was poor (1) at the back of the hall where the first balcony is too low above the heads of the audience (the sound arriving here has undergone excessive attenuation caused by the audience in front), (2) on the third balcony, where the ratio of direct to reverberating sound is too low on account of the distance from the stage, and (3) in many other places on those occasions when there is a great deal of audience noise (at revues and the like).

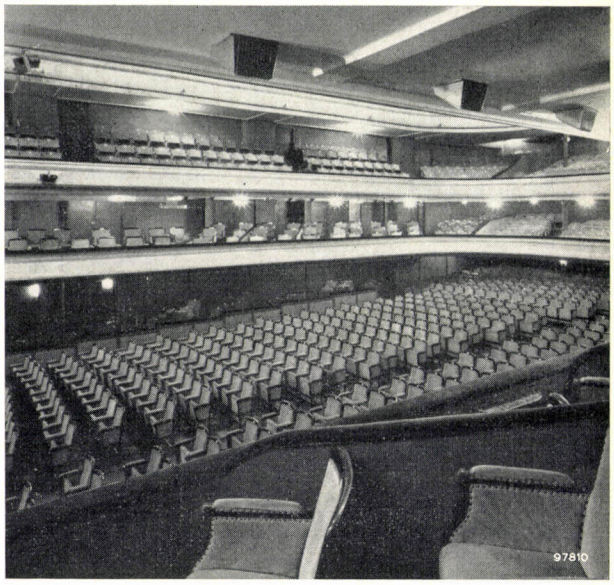


Fig. 15. Auditorium of the Gebouw voor Kunsten en Wetenschappen in The Hague, viewed from a box on the first balcony. Three of the small loudspeakers may be seen along the lower edge of the second balcony.

To improve intelligibility, four loudspeaker columns are mounted on either side of the front part of the auditorium, one pair at head height below the first balcony, another pair above that balcony, another above the second and another above the third balconies. These columns, each of which contains five EL 7021 loudspeakers, give coverage to the stalls and to side and rear balcony seats. However, sound from the columns is no better able than voices from the stage to penetrate the deep space under the first, overhanging, balcony. To reach seats here, the loudspeaker signal would have to be amplified so much that the volume would be excessive for members of the audience seated nearer to the columns, and there would be a risk of acoustic feedback into the microphones on the stage.

This being so, it was decided to make use of the delay equipment and supply the back stalls, via loudspeakers in the rear of the hall, with sound arriving about 10 milliseconds later than the sound from the stage and the columns. The impression given is that of sound coming from the stage. Only when one is closer than half a metre to the loudspeakers beneath the balcony is one conscious of the fact that sound is coming from them; however, they are mounted a good metre away from the nearest seats.

The same method has been employed to improve intelligibility in the back seats of the first, second and third balconies.

Thus the delay equipment can be operated in two ways:

(1) to produce an ambiophonic effect; all the loudspeakers around the auditorium (except those of the third balcony) are distributed amongst the four channels and the feedback is put into action;
(2) for making speech more intelligible: only the 16 + 16 + 12 + 10 loudspeakers referred to above are operative, being supplied from one of the delayed signal channels; the feedback is switched off.

Operation

First and foremost, the new installation had to be simple to operate; a second requirement was that it should be possible to make readjustments from one given place in the hall.

The first requirement was met by abandoning the idea of a large central control desk, the amplifier rack merely being provided with four switches marked "stage", "soloist", "orchestra" and "organ". Each switch brings into action a set of microphones appropriate to the type of performance indicated: the pre-amplifiers associated with the microphones are automatically switched on and their outputs connected to certain power amplifiers or to the ambiophony installation, the signals passing through attenuators that have been adjusted once and for all on the basis of listening tests. Switching is also possible for combinations such as orchestra and soloist, orchestra and organ, and stage with commentator (as in Arthur Miller's play "A view from the bridge").

The switch marked "stage" brings into action four EL 6031 microphones (*fig. 16*) mounted along the edge of the stage. The loudspeaker columns also come into action, and relays switch the delay equipment over to "speech" irrespective of whether other switches are on or not. In all other cases the delay equipment functions as part of the ambiophony installation. The switch marked "soloist" puts the soloist's microphone into operation, together with the columns and the ambiophony installation; the ratio between the intensities of the direct and indirect sound is adjusted in advance. The "orchestra" switch brings in the ambiophony installation only, the signals being derived from microphones above the podium. The "organ" switch links the organ microphone to the columns (which "put the organ on stage" acoustically) and to the ambiophony installation, which adapts the acoustics of the hall to organ music; in addition, low organ notes are backed up by loudspeakers in the ceiling.

The requirement that correction should be possible from some point within the hall has been met by providing a small mobile correction desk. Remote control is achieved very simply by making use of the new cadmium-sulphide photo-resistors [15]). Inserted in series with the volume controls on the amplifier rack, the CdS resistors are exposed to the light from small electric bulbs. The (direct) current flowing through each bulb can be controlled with the aid of a rheostat on the correction desk. The resistance of the cadmium sulphide, and hence the signal strength, varies with the luminous flux from the bulb. The system is quite free from crackle and is cheaper than a system with conventional crackle-free potentiometers. A second advantage is that all the cables between the correction desk and the amplifier rack carry direct current. There was therefore no need for the long screened low-impedance cables that had to be used in the Scala.

Fig. 16. Type EL 6031 electro-dynamic microphone on flexible mounting. The microphone has a hypercardioid directional characteristic.

It is possible from the correction desk to switch off various loudspeaker groups, singly or all together, by means of relays. Provided the "stage" switch is not on, it is also possible from this point to switch the delay equipment from "ambiophony" to "speech" and *vice versa;* one case in which this may be necessary is during operettas, when singing is interspersed with spoken passages.

[15]) N. A. de Gier, W. van Gool and J. G. van Santen, Philips tech. Rev. **20**, 277-287, 1958/59 (No. 10).

Volkswagen factory, Wolfsburg

We shall now leave the world of music and the theatre and give a short description of the unusual public-address installation in the Volkswagen factory.

It is laid down by law in the German Federal Republic that factory managements must call all employees together every three months and give them an account of the business and production situations. The Volkswagen factory employs about 35 000 people. The only meeting place capable of accommodating such a multitude is obviously the factory itself. Accordingly, the meetings are held in three or four adjacent aisles of the factory — *which* aisles are chosen depends on the circumstances at the time; alternatively, the meetings may take place in the open if the weather is good.

Each aisle is 250 m long, 25 m wide and only 8 m high. The roof rests upon heavy pillars spaced at intervals of 8 m and forming the divisions between the aisles. The surfaces of walls, roof, floor, pillars, machines etc. are very hard, acoustically, and do not by any means favour intelligibility. Equally unfavourable is the sawtooth shape of the roof, with its numerous, almost completely undamped, resonant cavities. Add to this the enormous length of the aisles (from the speaker to the furthest listener is about 200 m) and their lack of height, and it will be plain that achieving a fair degree of intelligibility was no easy problem. The problem was further complicated by the need for the electro-acoustical installation to be easy to set up and dismantle (this on account of the changing locale of the meetings), which meant that excessive decentralization had to be avoided; moreover, the installation had to ensure proper intelligibility in the open air as well as inside.

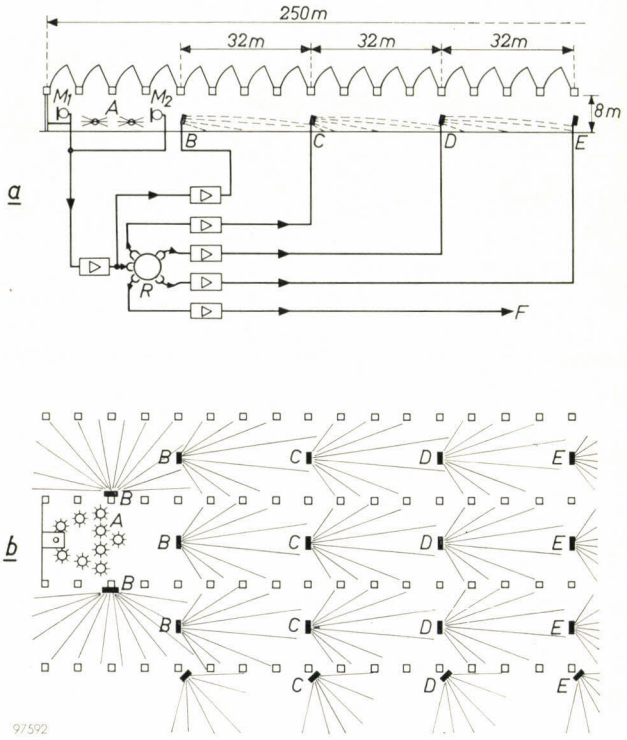


Fig. 17. *a*) Vertical section, and *b*) ground plan of a shop in the Volkswagen factory, where meetings attended by labour and management are held every three months. M_1, M_2: microphones. *A*: small loudspeakers with omnidirectional radiation pattern (see fig. 19). *R*: delay equipment supplying loudspeaker columns *B* . . . . *F* (*C* is delayed 0.1 s with respect to *B*, *D* is delayed 0.1 s with respect to *C*, and so on). The angles at which loudspeakers are mounted in their columns are staggered in such a way that sound is evenly distributed over the areas covered by the columns.

Fig. 18. Shop in Volkswagen factory during a works meeting. Some of the loudspeaker columns (*C*, *D*, *E* in fig. 17) may be seen in the photograph.

The design adopted was based on the principle illustrated in *fig. 17* (cf. fig. 17 of Part I). 32-metre sections of each aisle are supplied with sound from a loudspeaker column, the signal to each column being delayed by 0.1 s with respect to the signal fed to the column in front of it (this is roughly the time sound requires to travel 32 m). The columns have a slight forward tilt (fig. 17*a*), and the loudspeakers making up each column have a lateral skew such that they distribute sound evenly over the section of aisle to be covered (fig. 17*b*). *Fig. 18* is a photograph taken at one of the works meetings.

The above system has the advantage of giving hearers the impression that the sound is coming from the direction of the speaker. If it seemed to come from any other direction, their attention would soon start to wander, the tendency to embark on private conversations would increase, and the resulting increase in the noise level would mean decreased intelligibility, even less attention to the speaker, and so on.

A different solution has been adopted for the part of the factory (to the left in fig. 17) occupied by the speaker's chair and by tables at which members of the management and works council and representatives of the press are seated. Small loudspeakers with an all-round radiation pattern are set up on these tables (*fig. 19*). In view of the short range to the listener, the loudspeakers need only supply a small volume of sound. Consequently there is absolutely no danger of howl due to acoustic feedback.

The amplifiers and the delay equipment are mounted on a trolley in view of the desire for mobility. The setting up and dismantling of the whole installation do not take more than two or three hours.

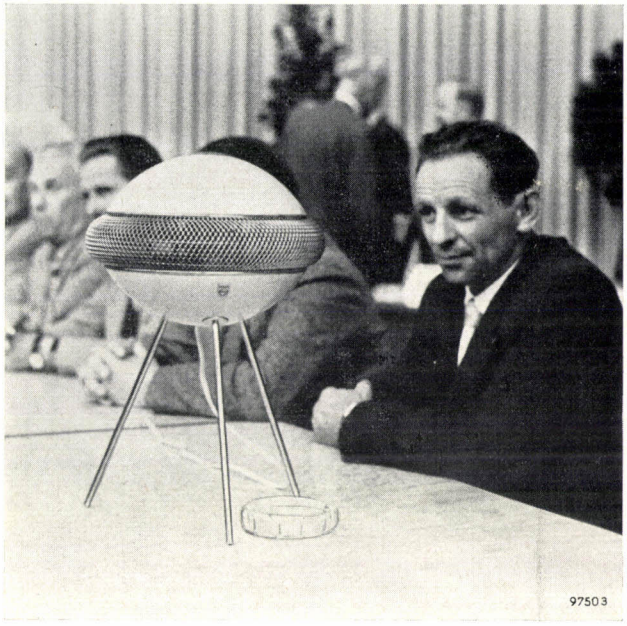


Fig. 19. One of the small omnidirectional loudspeakers (*A* in fig. 17) that are placed on tables occupied by management, works council and press.

For open air meetings the loudspeaker columns are grouped together in larger units. The signals to all units are in phase, the delay equipment being put out of action by means of a switch. In this manner speech can be reproduced with proper intelligibility in the open air.

Summary. The most important acoustical problems arising in theatres are those relating to intelligibility, to the acoustics for music played in the orchestra pit or elsewhere, and to the acoustics as they affect the actors and musicians. Visual requirements are often found to conflict with acoustic ones. The latter can however be satisfied by taking suitable steps to support speech and music with direct and indirect sound from an electro-acoustical installation. Requiring no special staff to operate it, an installation of this kind enables theatre premises to be exploited in new ways.

Besides an installation having purely artistic functions theatres have need of electro-acoustical facilities of other kinds, such as a monitoring system for management and artistes, a paging system for communicating with artistes in their dressing rooms, technicians and members of the public, an installation for the hard of hearing, a simultaneous-interpreting installation for use at international congresses, and so on. Closed-circuit television has also proved of value; one of its uses is to make the orchestra conductor visible to musicians in a separate studio, and another is to allow latecomers to follow stage proceedings from a foyer.

The sound control room is the heart of the whole electro-acoustical complex. Having described the equipment to be found therein, the author deals with the large installations that have been built in the Scala in Milan, the Théâtre National Populaire (Palais de Chaillot) in Paris, the Grand Auditorium at the 1958 Brussels World Fair, the Gebouw voor Kunsten en Wetenschappen in the Hague and the Volkswagen factory at Wolfsburg.

VOLUME 21, 1959/60, No. 3 pp. 73-108 Published 31st December 1959

Philips Technical Review

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

MOTION-PICTURE PROJECTION WITH A PULSED LIGHT SOURCE

by P. HOEKSTRA and C. MEYER. 778.554.4:621.327.534.3

Present-day motion-picture projectors are still almost invariably equipped with a continuously burning light-source — usually a carbon arc — the light from which is interrupted twice per frame period by a rotary shutter in order to obtain a stationary picture free from flicker. This article describes an illumination system in which the light source is a super-high-pressure mercury-vapour lamp operated by current pulses, and which dispenses with the need for a shutter.

As long ago as 1938 Philips introduced a motion-picture projector in which the light source was a discharge lamp, viz. a water-cooled super-high-pressure mercury-vapour lamp [1]). This lamp, which was fed by direct current and consumed 1000 watts (2 A, 500 V), replaced the conventional carbon arc. Both types of light source combine small dimensions of the radiant element with great brightness, a characteristic which is essential for application in film projectors [2]). All cinema-goers know that satisfactory projection can be achieved with the carbon arc. Nevertheless, this light source has certain serious drawbacks. In the first place, it requires constant attention from the projectionist [3]). The arc must be allowed to burn for several minutes before use, and the carbon electrodes must be long enough to last for the duration of the film. The incandescent ball of gas in the crater of the positive electrode, which is the main source of light emission, shows a tendency to shift its position with respect to the mirror focus. If the projectionist does not intervene, the result is a noticeable diminution of luminous flux on the screen and a change in the colour of the light. Furthermore, an exhaust system is required for removing the combustion products, but even this cannot prevent deposits forming at undesired places. The carbon arc also causes considerable heating of the film, at least of the picture areas; the edges remain cold. As a result the film buckles slightly, an effect which is accentuated by the fact that it is chiefly the photographic emulsion that absorbs the heat, and not so much the transparent base. This buckling adversely affects the sharpness of the picture — a drawback which is more noticeable nowadays than formerly. There are two reasons for this. Firstly, present-day films are generally made of non-inflammable material, and this buckles more than the material previously used. Secondly, there is a tendency to use faster projection lenses, which have less depth of focus. The fact that the film buckles is therefore more serious and it is more difficult to obtain a picture of uniform definition on the screen.

Compared with the carbon arc the super-pressure mercury lamp has appreciable advantages. It requires much less attention from the projectionist; in fact, without much trouble its operation can be made largely automatic. With the present shortage of skilled personnel, this is an attractive feature. A discharge lamp of this kind produces no unwanted deposits, thus dispensing with the need for an exhaust system, and it heats the film much less then a carbon arc with the same luminous flux, for one reason because the lamp produces much less infrared radiation. The lamp house, too, is much smaller than that for a carbon arc. Water-cooling is required, but this complication is more than offset by the advantages mentioned. In any case, the

[1]) Described by G. Heller in Philips tech. Rev. 4, 2-9, 1939.
[2]) See e.g. T. J. J. A. Manders, Incandescent lamps for film projection, Philips tech Rev. 8, 72-81, 1946, and also the article cited in footnote [1]).
[3]) See, for example, Mitchell's manual of practical projection, International Projectionist Publishing Co., New York 1956, pp. 112-117 and pp. 177-206.

use of high amperage carbon arcs also calls for water-cooling of the film gate. Still higher currents require that the positive carbon contact, too, be cooled.

The fact that the mercury lamp was unsuccessful in its first attempt to rival the carbon arc was mainly due to the deficiencies at that time in its luminous flux, in its rendering of red colours and in its reliability. The life of a mercury lamp ends fairly suddenly, and just when it will end cannot be predicted with any certainty. Consequently the lamp could not be changed in good time and its possible failure during a performance was something that had to be accepted. With the projector of 1938 it took some tens of seconds before the stand-by lamp came into operation. In the following we shall describe a new film illumination system using a super-pressure mercury lamp which overcomes these drawbacks whilst preserving the favourable features. The new system in fact possesses certain important additional advantages, including a very flicker-free picture and an improved efficiency. Some particulars of the lamp itself will also be mentioned. The illumination system can be incorporated into a projector of a new design, discussed elsewhere in this issue [4]). A photograph of this projector, type FP 20 S, equipped with the new illumination system, is shown in *fig. 1*.

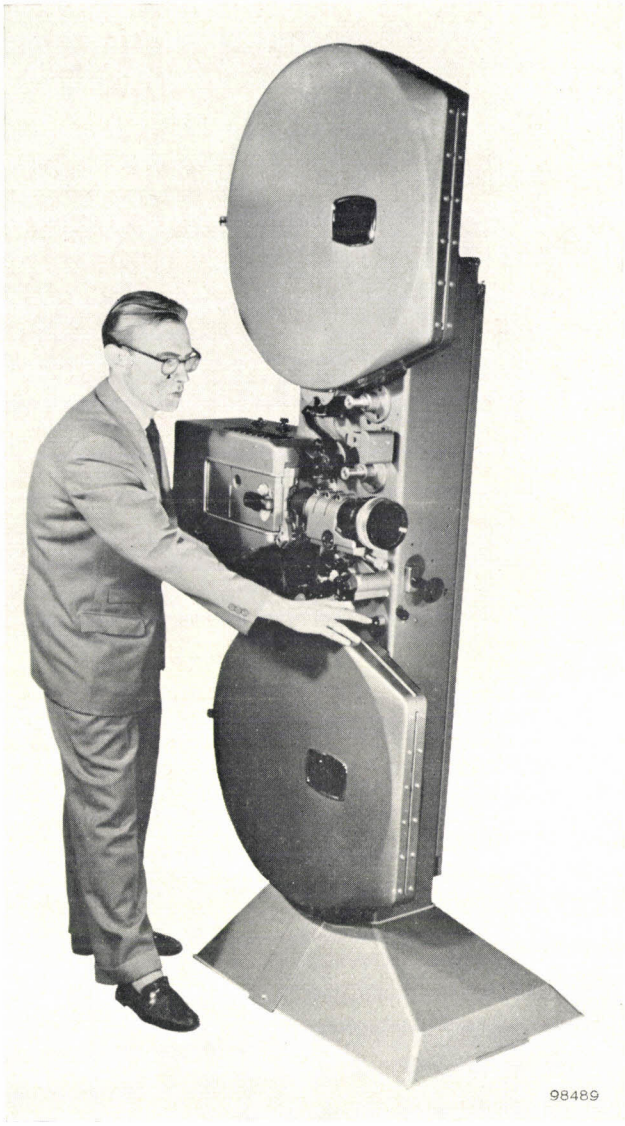


Fig. 1. The new Philips motion-picture projector, type FP 20 S, provided with the new illumination system which is the subject of this article. The projector itself is discussed elsewhere in this issue [4]).

The illumination system

As mentioned, the illumination system contains a water-cooled super-pressure mercury lamp as light source. This discharge lamp consists of a small tube of fused silica (length 8 cm, max. diameter 5.6 mm) in which a linear discharge is maintained between the electrodes (spacing 17 mm). The light output of the lamp increases more than proportionally with the power consumed, since the efficiency of high-pressure mercury lamps increases with the power supplied per unit length of discharge [5]). Higher loading improves the colour rendering at the same time, the reason being that the share of the continuum in the emitted light increases with respect to that of the line spectrum. Unfortunately the life of the lamp is thereby shortened, mainly because of the higher temperature of the fused-silica wall. In the projector brought out in 1938 this set a limit to the quantity and the quality of the light output. At the time that this projector was developed an attempt was already made to solve the heat problem by loading the lamp intermittently instead of continously. In motion-picture projection this is an attractive proposition in as much as the illumination of the film must in any case be interrupted twice during each frame period: once in order to pull down the film from one frame to the next without this being visible on the screen, and a second time in order to increase the number of light pulses per second sufficiently to prevent the picture from flickering [6]). The energy consumed by the lamp during the interruptions yields no useful light. If the lamp is loaded only when the generated light is usefully employed, the instantaneous load can be raised whilst maintaining the same

[4]) See the following article in this issue: J. J. Kotte, A motion-picture projector of simplified design, Philips tech. Rev. **21**, 83-77, 1959/60.

[5]) See e.g. W. Elenbaas, Fifty years of the high-pressure mercury vapour lamp, Philips tech. Rev. **18**, 167-172, 1956/57.

[6]) For a detailed discussion of this phenomenon, see: J. Haantjes and F. W. de Vrijer, Flicker in television pictures, Philips tech. Rev. **13**, 55-60, 1951/52.

mean load. The result is more useful light, and moreover the changed spectral distribution of the light provides better colour rendering. The luminous intensity of a mercury lamp follows current variations virtually without inertia. When fed with current pulses, then, the lamp flashes on and off synchronously. This makes it possible to dispense with the rotating shutter which, in the conventional continuously-burning carbon arc (or other source), intercepts the light twice per frame.

Tests with a pulsed mercury lamp were initially disappointing. The life of the lamp was found to decrease sharply when the instantaneous load was raised, even though the mean load remained unchanged. Only after a great deal of development work on the lamp, and careful matching of the pulse generator to the lamp, was it possible to achieve an acceptable life with an instantaneous load high enough to ensure good colour rendering. The final version of the lamp represents a compromise between useful life on the one hand and light output (determined by the *mean* load) and colour rendering (determined by the peak amplitude of the current pulses) on the other. The choice fell on a mean load of 800 W with current pulses of about 15 A peak amplitude. The mean life is then 33 hours. The luminous flux incident on the screen is equal to that from a 60 A carbon arc, and is therefore adequate for the majority of cinemas; the colour rendering, too, is entirely satisfactory. The lamp has been given the type designation SPP 800 W.

Increased projection frequency

We have already mentioned as a particular advantage the very flicker-free picture obtained. This is due to the light source flashing *three* times per frame (*fig. 2*). The frequency is thus 72 flashes per second instead of the usual 48 (the number of frames per second remains unchanged of course). The sensitivity of the eye to flicker increases with the brightness of the picture, and moreover the periphery of the retina is more sensitive to it than the central part. The lack of flicker in the picture from the new projector is therefore particularly noticeable in the brighter projection of wide-screen pictures, such as "CinemaScope."

The higher frequency of 72 flashes per second is in principle also possible with a continuously burning light-source if a three-bladed shutter is used. The resultant loss of light, however, would be unacceptable. With a film-transport mechanism employing the usual four-slot Maltese cross, the shutter would then intercept 75% instead of 50% of the light output (see *fig. 3*).

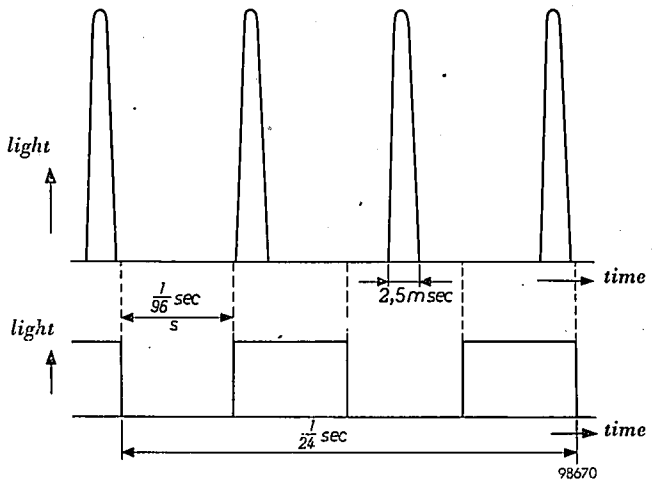


Fig. 2. Distribution over one frame period (1/24 sec) of the three light pulses from the mercury lamp type SPP 800 W (above), and (below) distribution of light periods over a frame period in a normal projection system with shutter. The film is transported during the interval denoted by *s*.

Lamp holder and turret

Fig. 4 shows a dismantled lamp holder together with a type SPP 800 W lamp. The components are so made that they mate together correctly without any adjustments during assembly. The round glass window in front of the lamp is a filter which absorbs ultraviolet radiation. The cooling water that flows over the lamp also passes over the filter, thus dissipating the heat generated in it. Of the little infra-red radiation emitted by the lamp the major part is absorbed by the coolant. The heating of film and film gate is therefore slight.

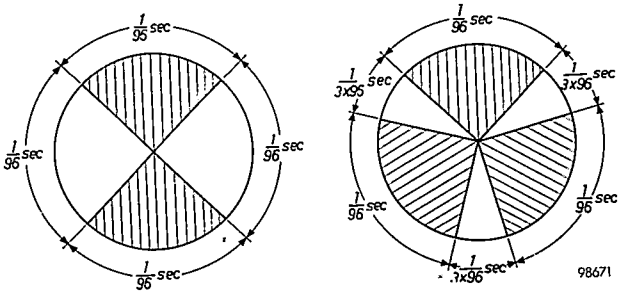


Fig. 3. The theoretical loss of light caused by a shutter that completes one revolution per frame period is equal to the fraction of 360° covered by the blades jointly. Since the fundamental frequency of the illumination cycle is mainly responsible for the nuisance of flicker, the shutter blades must be identical; otherwise the fundamental frequency of the illumination cycle would be equal to the frame frequency (24 per second). The film-transport mechanism is almost invariably a normal four-slot Maltese cross [7]). The frame-shift period must then be $\frac{1}{4}$ of the total period per frame, so that each shutter blade must cover 90°. With a two-bladed shutter (left) the theoretical light-loss is thus 50%, and with a three-bladed shutter (right) 75%.

[7]) The Maltese cross is described in J. J. Kotte, A professional cine projector for 16 mm film, Philips tech. Rev. **16**, 158-171, 1954/55, in particular fig. 7.

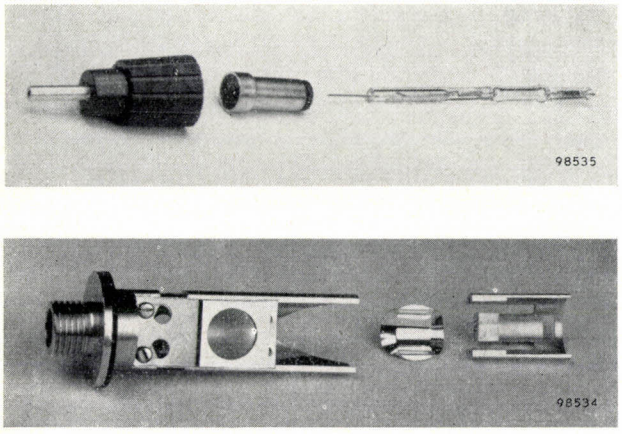


Fig. 4. Dismantled lamp holder. From left to right, above: the head of insulating material with silver contact peg, the centring bush (also acting as coolant seal) and the lamp. Below: the holder with filter for absorbing ultra-violet radiation, the cylindrical reflector, and its holder. The assembled lamp holder is shown in fig. 5.

Fig. 5 shows how two lamp holders are fitted in a "lamp turret". The lower holder, shown removed from the turret, contains the lamp normally in use, the upper holder contains the stand-by lamp. If the lamp in the lower holder becomes defective, the upper one automatically takes its place, the lamp turret pivoting about a horizontal spindle under its own weight. This is effected by the lower lamp being kept in its operating position by a pawl which is tripped when the lamp current cuts out. The cooling water flows in along the spindle around which the lamp turret rotates, and reaches the operative lamp holder via corresponding holes in the spindle and the turret. The cooling water is thus automatically switched to the stand-by lamp as soon as this is called into operation. The same applies to the electric supply. The stand-by lamp takes over so quickly as to be imperceptible to the audience watching the screen. The projectionist notices it, however, from the colour of a transparent sector on the front of the turret. If the projector is operating normally the green sector is visible; when the stand-by lamp enters into operation the red sector appears as a warning to the projectionist. He can now unscrew the lower lamp holder from the turret and replace the lamp whilst the projector remains in operation. Later, when there is a suitable opportunity, he can turn the lamp turret by hand back to its normal position. In this way the useful life of each lamp is exploited to the full, and the performance is insured against the unlikely event of a double lamp failure.

The above implies that, as far as the lamp is concerned, there is no limit to the lengths of film that can be used. This advantage promises to gain in importance in the future. In connection with the risk of fire, film spools and drums were formerly designed to accommodate films not longer than 600 metres. Now that inflammable films are no longer commonly used and have in fact been prohibited in many countries, it is likely that longer lengths of film (up to 1800 metres) will come increasingly into use.

When a lamp holder and its lamp are loaded into the turret, the exposed end of the lamp (see lamp-holder in fig. 5) slides into a hole surrounded by an annular nozzle inside the turret. The electrode pin at this end then presses against a spring-loaded contact, which also acts as a valve shutting off the water supply to the nozzle when no lamp is mounted. The other lamp electrode pin is clamped inside the contact peg at the head of the lamp holder (fig. 4, upper photo). When the lamp holder is in the operating position, this peg lies on a contact consisting of a metal flat partly encapsulated in a block of insulating material. This block is fixed to the lamp house in which the lamp turret is incorporated (*fig. 6*).

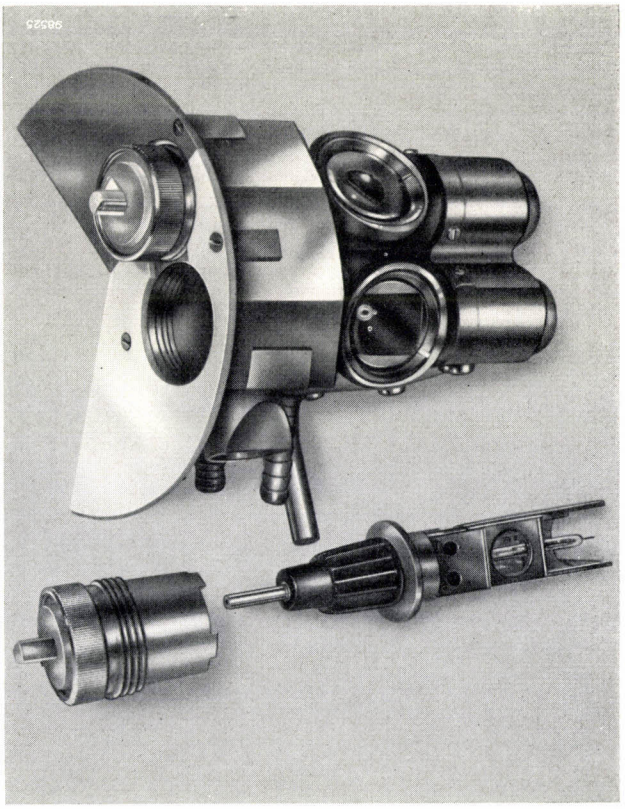


Fig. 5. Lamp turret with two lamp holders. The lower one, which is removed from the turret, contains the lamp normally in use; the upper lamp holder contains the stand-by lamp. The windows in front of the lamps also act as the first condenser lenses. The turret is provided with a red and a green sector, which indicate to the projectionist at a glance which lamp is in operation.

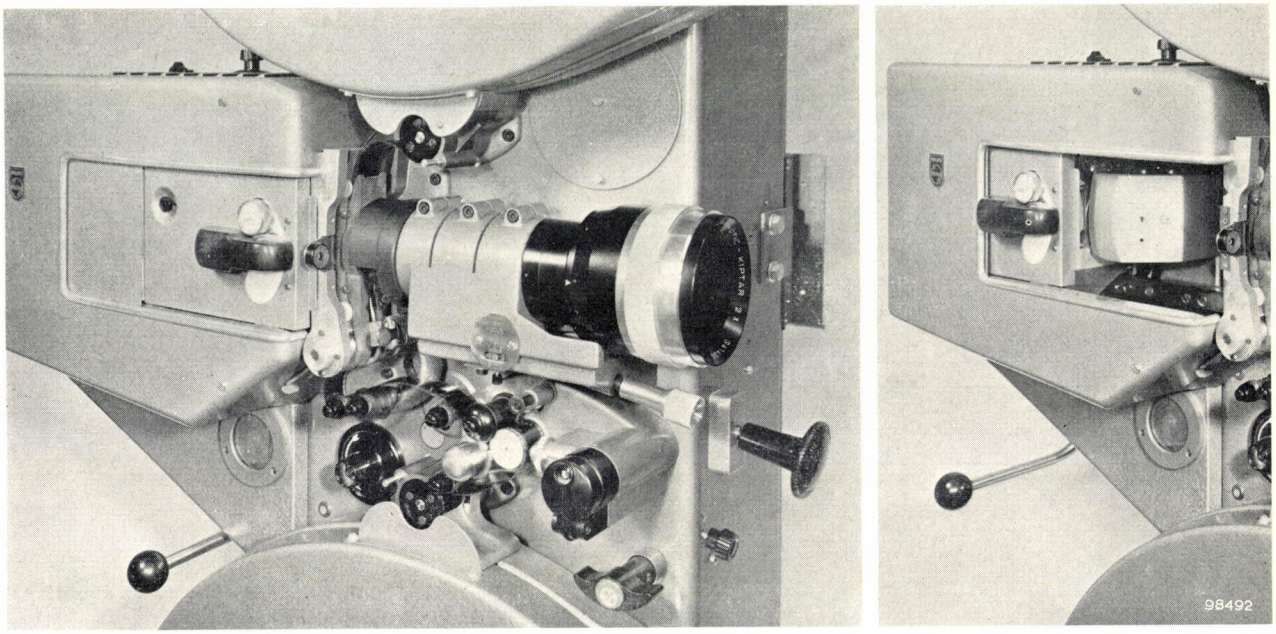


Fig. 6. Part of cine-projector fitted with the pulsed light-source illumination system. Above the block of black insulating material, which carries the positive contact, can be seen the cap which encloses the lamp holder with the stand-by lamp, in the turret. The photograph shows on the left the lamp house in the position for film projection. By means of the lever (bottom left) the lamp house can be moved backwards. The same lamp then serves for projecting lantern slides (right).

The optical system

The optical system is shown schematically in *fig. 7*. Seen from the light source, the film and aperture are located directly behind the condenser lenses. The effective diameter of the condenser is therefore only slightly larger than the diagonal of a film frame (26 mm in standard films). Since this implies a small condenser, the light source must be close to it in order to ensure that the condenser receives a sufficient part of the luminous flux. In this way a very compact assembly is obtained. The fact that it is technically feasible is due in the first place to the small dimensions of the lamp and in the second place to the simple way in which the ultraviolet and infra-red radiation is intercepted and rendered harmless. This radiation would otherwise be largely absorbed by the condenser and would damage it. Precisely because it is so small, the condenser is in any case severely heated merely by the high flux of visible light, of which it absorbs a very small fraction [8]).

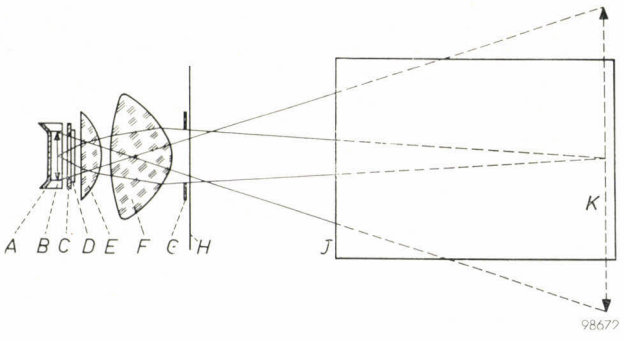


Fig. 7. The optical system. *A* cylindrical mirror, *B* linear light-source, *C* diaphragm, *D* ultraviolet-absorbent filter, *E* first condenser lens, *F* second condenser lens, *G* film aperture, *H* film, *J* projection lens, *K* image of light source projected by condenser.

The condenser system consists of two lenses, the first of which also acts as a water-tight window for the lamp turret. Behind the light source there is a cylindrical reflector of aluminium sheet (0.3 mm thick). Since this mirror provides more than 60% of the projected light, it is essential that its reflective power should not be diminished. For this reason a new reflector is supplied with every new lamp.

In a projection system as in fig. 7 the condenser is always arranged to form an image of the light source at the position of the projection lens. It is not the image formation itself that is important (the image is usually unsharp) but the fact that a pencil of light emanating from an arbitrary point of the light source has a small cross-section *at* the projection lens. This being so, if the central ray of

[8]) With conventional light-sources a different and less compact system is employed for film projection. The two systems are compared in the article mentioned in footnote [1]) and more recently in: P. M. van Alphen and M. Bierman, A mirror condenser lamp for 8 mm projectors, Philips tech. Rev. **19**, 233-235, 1957/58.

a pencil from any point of the source passes through the projection lens, all or almost all of that pencil of light does so. The optical centre of the condenser falls virtually in the film and hence virtually in the focal point of the projection lens (the screen is a considerable distance away). Consequently, the relative aperture of the projection lens determines what length of the light source is effective in contributing to the projected light. In the pulse-operated projector the central rays emanating from the ends of the discharge make such large angles with the axis that the projection lens should have at least a relative aperture of f 1.4 in order to transmit them. A normal value for the relative aperture of a projection lens is f 2. If such a projection lens is used, the extremities of the light source do not contribute to the luminous flux on the screen. This is done deliberately, for it precludes that the luminous flux incident on the screen gradually diminishes with the age of the lamp, as a result of the darkening of the fused-silica wall (a re-crystallization or devitrification phenomenon) which begins at the extremities of the discharge.

To avoid needless heating of condenser and film gate, a diaphragm is interposed between the lamp and the condenser. This is matched to the projection lens and intercepts the unused light. This diaphragm also serves as holder for the ultraviolet filter; it can be seen in fig. 4.

Projection of lantern slides

The same illumination system is employed for the projection of lantern slides. The slide holder is located at the other side of the projector. In order to convey the light to that side, the lamp house is mounted on rails. By means of a lever (see fig. 6) the lamp house can be shifted backwards. A plane mirror is thereby moved into the path of the rays at an angle of 45° (right, fig. 6). (When the lamp house is returned into position for film projection, this mirror folds upwards again.) The light reaches the lantern slide via the lens E (*fig. 8*), a second 45° mirror F and the lens G, whence it is transmitted to the lantern slide projection lens. In *fig. 9* the lamp house is shown in an intermediate position, in which the mirror has not yet descended.

The lamp is also pulsed for the projection of lantern slides, in order to benefit from the improved colour rendering resulting from pulsed operation.

Luminous flux and efficiency; light distribution

With a standard film gate, but with no film in it, and using a projection lens of the Petzval type

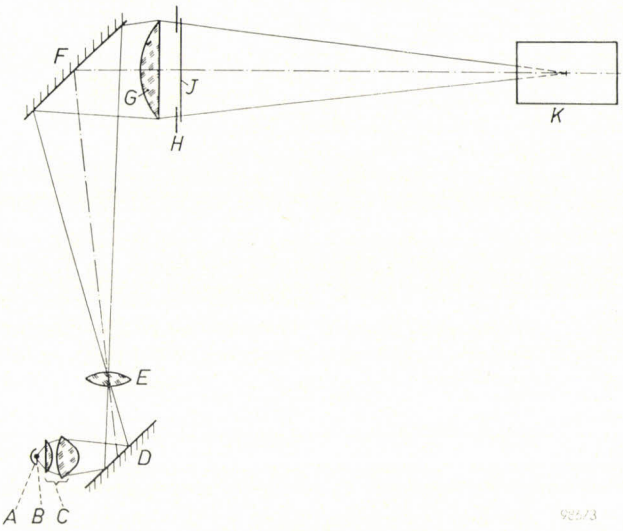


Fig. 8. Path of rays in the projection of lantern slides. A cylindrical mirror, B light source, C condenser, D plane mirror, E lens, F plane mirror, G lens, H mask, J lantern slide, K projection lens for lantern slides.

having a relative aperture of f 2, and a pulsed light-source, operated at 800 W, the system delivers a luminous flux on the screen of 4800 lumens. With a projection lens of the same type, but having a relative aperture of f 1.6, the luminous flux on the screen is as high as 6000 lumens [9]). The efficiency in these two cases is thus 6 and 7.5 lumens per watt, respectively. The earlier-mentioned 1938 projector, with a lens of relative aperture f 2 and

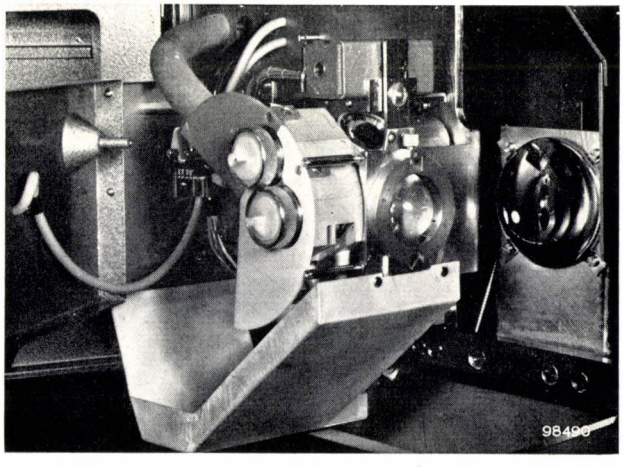


Fig. 9. Lamp-house opened to position between that for film projection and that for lantern slide projection. The 45° mirror has not yet descended into the path of the rays. Below can be seen the tray for collecting any leakage water. The cooling water is fed in through the thick hose visible above.

[9]) A projection lens system of the Petzval type consists of four lenses. Systems having a relative aperture of f 1.6 are usually more complicated and contain six or seven lenses; the gain due to the greater relative aperture is then partly lost as a result of higher losses in the lens system.

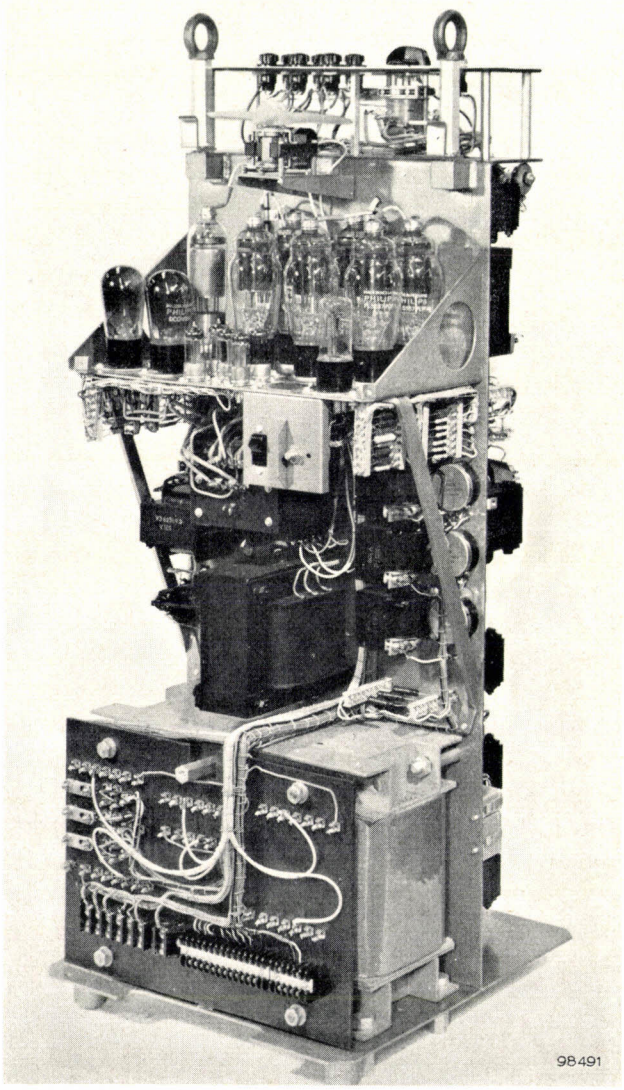


Fig. 10. Power pack (pulsator) for the pulsed-source illumination system, with protective hood removed.

a rotary shutter, gave a total flux of 2900 lumens [10]). The lamp consumed 1000 W, so that the efficiency was 2.9 lm/W. The more than double efficiency of the pulsed illumination system is, of course, mainly attributable to the fact that light is only produced when it is usefully employed. For carbon arcs of 45 and 60 amperes the flux on the screen is respectively 3500 and 5500 lumens. The efficiency of a carbon arc is approximately 2.5 lm/W.

According to the practice in projection technique, the fluxes mentioned are based on measurements of the luminous intensity in the middle of the screen. Since the luminous intensity always decreases from the middle towards the edges, the actual flux values are lower. In the case of the pulsed lamp the decrease towards the edges amounts at the most to 10% in the horizontal direction and to between 15 and 20% in the vertical direction. (The smaller horizontal decrease is due to the elongated form of the horizontally positioned lamp.) This is appreciably less than in the case of the carbon arc, where the decrease for the same flux values is 25% horizontally and about 20% vertically [11]). For the same values of luminous flux, as given according to the normal practice, the true luminous flux of the pulsed system here described is therefore actually greater than that of a carbon arc.

Power pack (pulsator)

The power pack which supplies the current pulses for the lamp (the "pulsator") is shown in *fig. 10*. In the simplified block diagram, given in *fig. 11*, *A* is a six-phase rectifier which charges the capacitor *C* to a voltage higher than the operating voltage of the lamp. At the moments when the lamp is required to flash, a voltage pulse is applied to the grid of thyratron *T*, making the latter conduct. The capacitor *C* thereupon discharges through the lamp.

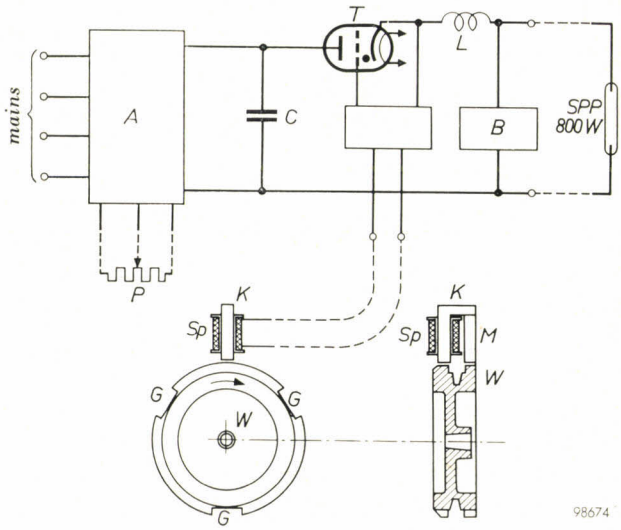


Fig. 11. Simplified block diagram of power-supply system for the pulsed super-pressure mercury lamp, type SPP 800 W. *A* six-phase thyratron rectifier. *P* potentiometer for varying the voltage on the thyratron grids. *C* capacitor which supplies the current pulses. *T* thyratron to whose grid the synchronizing pulses are supplied which are generated in the coil *Sp*. This coil is wound on the soft-iron core *K* and is located in the projector. *W* flywheel of Maltese cross. The grooves *G* in *W* produce in *Sp* three voltage pulses per revolution of the flywheel. *L* inductance which gives the condenser discharge the nature of a damped oscillation, and which together with *C* governs the flash duration. *B* holding-current rectifier; this keeps a small current flowing through the lamp during the intervals when the lamp emits no light.

[10]) In the article cited in footnote [1]) a figure of 2500 lumens is given. This referred to a lens system that was not provided with the low-reflection coated lenses now commonly employed.

[11]) With a carbon arc the decrease towards the edges can certainly be reduced, but only at the expense of the illumination level in the middle of the screen. Some decrease is desirable, as otherwise contrast effects would make the edges of the screen appear brighter than the middle.

The inductance L then tends to produce an oscillation in the circuit. At the moment, however, that the current tries to change direction, the thyratron becomes non-conducting and the lamp is extinguished. The rectifier charges up the capacitor, and upon the arrival of the next voltage pulse on the grid of T the lamp flashes again. At the same time the lamp is fed by the "holding current" rectifier B. This ensures that in the period between the flashes a small current continues to flow through the lamp, thereby keeping the gas sufficiently ionized to ensure that it flashes readily upon the arrival of the succeeding pulse. (If the lamp is cold, the gas pressure is low and so therefore is the ignition potential; switching on the lamp when no holding current is yet flowing therefore presents no difficulty.) The holding current is very low and produces no light of any significance.

The duration of the flash depends on the period of the oscillation produced in the circuit, and this period can be regulated by varying L or C, or both.

Rectifier A operates with six thyratrons. The voltage delivered by the rectifier, and to which C is charged up, can be adjusted by means of a variable bias on the grids of these thyratrons. In this way it is possible to control the amplitude of the current pulses. The variation is effected with potentiometer P (fig. 11), which can either be mounted on the projector or placed at any other desired position, e.g. in the cinema auditorium. Flash duration and peak current together determine the energy per flash and also — at a given number of flashes per second — the average lamp-load. If this load varies from 600 to 800 W, the power pack consumes from 1.3 to 1.6 kW. A comparable carbon-arc installation consumes approximately 3 kW.

Synchronization and protection

The three light flashes per frame must, of course occur outside the pull-down period. The synchronizing pulses for the grid of the thyratron T are therefore derived from the flywheel of the Maltese cross in the projector. This steel flywheel, in which three grooves G are milled (fig. 11) forms part of a magnetic circuit containing, in addition, a permanent magnet M and a soft-iron core K around which a coil Sp is wound. Every time one of the grooves passes this assembly, the magnetic flux is sharply attenuated. The resultant voltage pulse in the coil Sp is applied to the grid of T. The exact moments at which the flashes occur can readily be adjusted by shifting the position of the magnet-core-coil assembly in relation to the flywheel. This can be done whilst the projector is operating.

Precautions are taken to ensure that the lamp, in the event of a fault, cannot burn continuously directly from the rectifier A, and also to prevent it operating only from the holding-current rectifier.

When the apparatus is switched over for lantern-slide projection, the projector is not running. The synchronizing pulses for the thyratron T are then derived directly from the mains. Pulses are then delivered with twice the mains frequency. On 50 c/s mains this means that there are 100 instead of 72 pulses per second as for film projection. To avoid overloading the lamp the capacitor voltage is then at the same time automatically reduced such that the average load remains the same.

Automatic change-over

A motion-picture film often consists of two or more reels, and for this reason most cinemas have two projectors installed. When one reel is at an end, the next reel, which is already loaded on the other projector, is switched on without interrupting the projection. A correct "change-over" calls for vigilance on the part of the projectionist. With a projector equipped with the new illumination system the change-over can easily be made automatic. For this purpose an automatic change-over device has been designed. It consists of a base plate on which several microswitches are mounted, and these are operated by cams on a common motor-driven shaft. When one reel is nearly at an end, a warning sign appears in the top right corner of the projection screen, which is the signal for the projectionist to start the motor for the other projector. To automatize the change-over a contact strip is affixed to the film, adjacent to the motor-starting sign, between the edge and the perforations. This contact strip actuates a relay which starts the idle projector turning and switches on at the same time the motor that drives the automatic change-over device. The cams then start to rotate, whereupon 1) picture and sound are switched over after 8 sec, 2) some seconds later the "run-off" projector is stopped and 3) the cam-drive motor is switched off. The automatic device is now ready for the next change-over cycle. This system relieves the projectionist of quite a number of routine operations.

The lamp

The D.C. lamp type SP 1000 W used in the 1938 projector was not the immediate starting point for the development of the SPP 800 W lamp used in the new illumination system. The SP 1000 W type contains a surplus of mercury, that is to say so much that, even when the lamp is burning, only a very

small fraction of the mercury is vaporized. This lamp has a number of drawbacks which become particularly apparent when used for film projection and which will be touched on presently [12]). A start was therefore made on the development of a so-called *dosed* D.C. lamp, that is to say a type containing so little mercury that practically all of it is vaporized when the lamp is burning (*fig. 12*). This work was subsequently diverted to the development of a dosed lamp for *pulsed* operation.

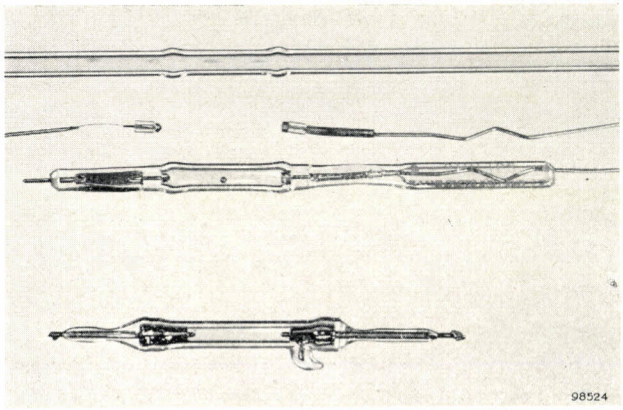


Fig. 12. The dosed super-pressure mercury lamp SPP 800 W for pulsed operation. The photograph shows from top to bottom: the pre-shaped tube of fused silica in which the electrode systems are sealed; the electrode assemblies; and the complete lamp. A fused silica capillary fits over each electrode stem. In the complete lamp the dosed quantity of mercury can be seen in the form of a droplet. Below, the surplus-mercury lamp type SP 1000 W. The mercury can be seen around the electrodes.

There are two chief disadvantages attaching to the lamp containing a surplus of mercury. Before igniting the lamp it is desirable to shake and tap it in order to distribute the mercury uniformly around the two electrodes which project some way into the discharge space. This may prove rather troublesome, particularly in a lamp that has already burnt for several hours. Furthermore, the lamp sometimes burns unsteadily owing to the abrupt rising or falling of the arc voltage. This phenomenon is probably connected with changes in the position of mercury surfaces.

The dosed lamp

A dosed lamp cannot be made by simply taking a lamp identical to the old surplus-mercury type and providing it with a dosed quantity of mercury. The electrodes would then project too far into the discharge space. The distance from the point of an electrode where the discharge is initiated — and where the heat is produced — to the seal would be so great, so that the temperature at this seal would be considerably below the temperature in the discharge. The latter temperature is limited by what the fused silica can stand, and the lowest temperature in the lamp (which determines the mercury vapour pressure) would therefore be lower than in the case when the surplus mercury *is* present. The mercury vapour pressure would therefore not reach the required high value. To produce a dosed lamp in which the mercury-vapour pressure rises just as high as in a surplus-mercury type, the end-wall must be located at approximately the position taken up by the mercury surfaces in the surplus-type lamp (see, fig. 12, below). This represents a problem, however. In the surplus-type lamp the electrodes are tungsten wires. Since these cannot be directly sealed in fused silica, the wires are first sealed in an intermediate glass which, in its turn, is sealed in the fused silica. This form of intermediate seal is not suitable for the dosed lamp because the spaces behind the ends of the electrodes cannot be made small enough, and moreover the intermediate glass seal cannot withstand the necessary high temperature. For this reason a molybdenum-foil seal was adopted, thereby eliminating the need for an intermediate glass. Very thin foils of molybdenum can quite readily be sealed in silica [13]). The foils used here are 10 to 12 microns thick. The lead-in wire is welded to one end of the foil and the tungsten wire electrode to the other end (fig. 12). However a foil seal is itself too weak to secure an electrode with sufficient mechanical strength. On each electrode wire, therefore, is a closely fitting capillary of fused silica. The whole electrode assemblies are then introduced into the fused silica lamp tube and sealed in. The Mo foils then fuse into the silica tube, and likewise the pieces of capillary. This leaves a tiny annular gap around each electrode wire, so narrow that the capilliary still firmly holds the wire.

Care must be taken that the temperature developed in the gaps round the electrodes is not below that of the wall near the discharge. With this in mind the diameters of the electrode wires are chosen so as to ensure that the lamp current generates the right amount of heat in the wires for this purpose. Another problem is the dissipation of heat. In the surplus-type lamp the mercury around the electrodes is responsible for considerable heat dissipation by conduction; in the dosed lamp the heat must be dissipated mainly by radiation. Each of the

[12]) The SP 1000 W lamp is well suited to other applications, for example in shipyards for apparatus in which the lamps are not switched on and off very frequently.

[13]) D. Gabor, D. R. P. 573 448, 1931. See also J. L. Ouweltjes, W. Elenbaas and K. R. Labberté, A new high-pressure mercury lamp with fluorescent bulb, Philips tech. Rev. **13**, 109-118, 1951/52, especially p. 115.

electrodes is therefore fitted at the end with a head to obtain a sufficiently large heat-radiating surface. The edges of the electrode heads are so close to the wall of the lamp that it is necessary to widen the bulb slightly at this position. Otherwise the bulb temperature here would become so high as to cause rapid devitrification of the silica. In fig. 12 it can be seen that the widening of the bulb at these positions is effected before the electrodes are sealed in.

As stated, the dosed lamp was developed with an eye on pulsed operation. DC pulses impose a heavier load on a discharge lamp than ordinary DC operation. The life of a lamp intended for DC operation is therefore almost invariably very much curtailed if the lamp is pulsed at the same mean power. One reason for this is the heat generated in the electrode seals. The heat generated at any given instant is, of course, proportional to the square of the current, and this means that, at the same mean power, the heat generated is greater the shorter the duration of the pulses. One cannot reverse the argument however, and conclude that a lamp which has a good useful life under pulsed operation will have a much longer life if operated on direct current. In order to obtain the most favourable conditions for DC operation various factors make it necessary to adopt, among other things, different dimensions for the electrodes.

One of the most important points in the development of the pulsed lamp was the control of manufacturing conditions. For motion-picture projection there is a constant demand for more light, and lamps for this purpose are always loaded to the utmost. This means that any manufacturing deviation of the lamp from the exact design almost invariably means a lamp of considerably shorter life. Mechanization of the manufacturing process was therefore just as necessary to obtain a uniform product as it was to lower the production costs. The average life of 33 hours mentioned at the beginning of this article applies under average conditions encountered in motion-picture projection, that is to say with the lamp being switched on or off every quarter of an hour (this period was determined by counting the number of occasions the lamp was actually switched on and off in an arbitrary cinema). In spite of all precautions, the spread in lifetimes is considerable. This presents no practical difficulties, however, because of the automatic change-over to the stand-by lamp.

At a mean load of 800 W the lamp gives a light output of about 40 000 lumens. The fact that the SPP 800 W lamp gives good colour rendering can really only be demonstrated by viewing a colour film projected with this light source. The improvement compared with the SP 1000 W is in fact much greater than might be expected from a comparison of the light-flux in spectral bands for the two types of lamp as shown in the *table*. (This spectral-band method was devised by Bouma [14]) who divided the whole visible region of wavelengths into eight sections, in a specified way, taking into account the properties of the eye.) It can be seen from the table that the SPP 800 W lamp, as compared with the SP 1000 W, shows appreciably less luminous output at the blue end of the spectrum, and a greater output at the red end.

Table. Relative luminous-flux of spectral bands for the lamps SPP 800 W and SP 1000 W. The visible spectrum is divided into eight bands after Bouma [14]). For comparison, the distributions are also given for daylight and for the high intensity carbon arc. The colours are indicated at the foot of the table.

Section	1	2	3	4	5	6	7	8
Wavelength range mμ	400-420	420-440	440-460	460-510	510-560	560-610	610-660	660-720
SPP 800 W	0.025	0.28	0.50	4.24	43.7	46.6	4.3	0.28
SP 1000 W	0.042	0.53	0.87	4.6	52.6	37.6	3.4	0.25
Daylight	0.025	0.26	0.91	11.1	40.8	36.2	9.9	0.73
Arc	0.050	0.27	0.97	10.2	43.7	33.2	10.6	0.94

436 495 566 589 627

violet blue green yellow orange red

[14]) See e.g. P. J. Bouma, Colour reproduction in the use of different sources of "white" light, Philips tech. Rev. **2**, 1-7, 1937.

Summary. The water-cooled super-pressure mercury lamp has many properties that make it more attractive as a light source for motion-picture projection than the conventional carbon-arc. In a normal projector, however, in which the shutter intercepts about 50% of the light, the light yield is rather on the low side for large theatres, whilst in colour projection less than justice is done to the red colours. This article describes an illumination system in which the mercury lamp is pulsed, thus producing light only when it can be usefully employed and rendering a shutter superfluous. At the same power consumption, this gives twice as much light on the screen as obtained from a continuously burning lamp, half of whose output is intercepted by a shutter. With a projection lens having a relative aperture of f 2, the luminous flux on the screen is 4800 lumens, and with a lens of relative aperture f 1.6, as much as 6000 lumens. The lamp consumes 800 W, so that the efficiency is respectively 6 and 7.5 lumens/watt on the screen, compared with 2.5 lumens per watt in the case of a carbon arc. Owing to the heavier instantaneous loading during the pulses the continuum in the spectrum is intensified with respect to the spectral lines, resulting in good colour rendering. Three light flashes of about 2.5 milliseconds duration are emitted per frame. This gives a strikingly flicker-free picture, even when the picture is very bright and wide. Apart from the illumination system proper, the power supply equipment is also discussed. The mercury lamp employed, type SPP 800 W, is a dosed type and was specially developed for pulsed operation in motion-picture projectors. Some particulars of its development are given.

A MOTION-PICTURE PROJECTOR OF SIMPLIFIED DESIGN

by J. J. KOTTE.

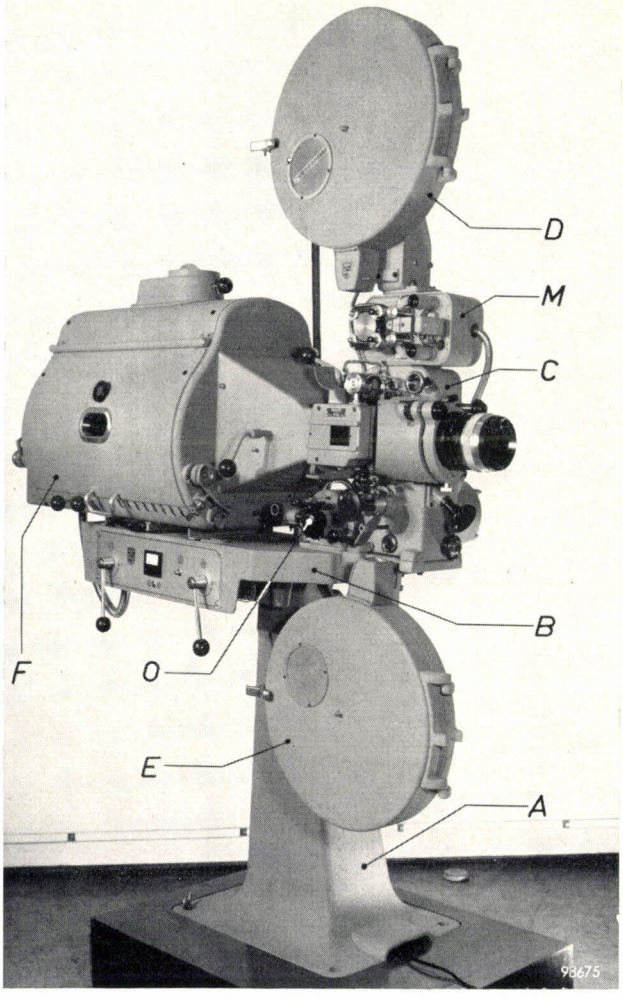
778.554.4:621.327.534.3

Fresh thought on the problem of how best to meet the demands of modern motion-picture projection has resulted in a projector of new design. It has proved possible to simplify the structure as a whole as well as various details, without, however, making concessions in regard to quality. The projector can be equipped with the pulsed light-source discussed in the preceding article.

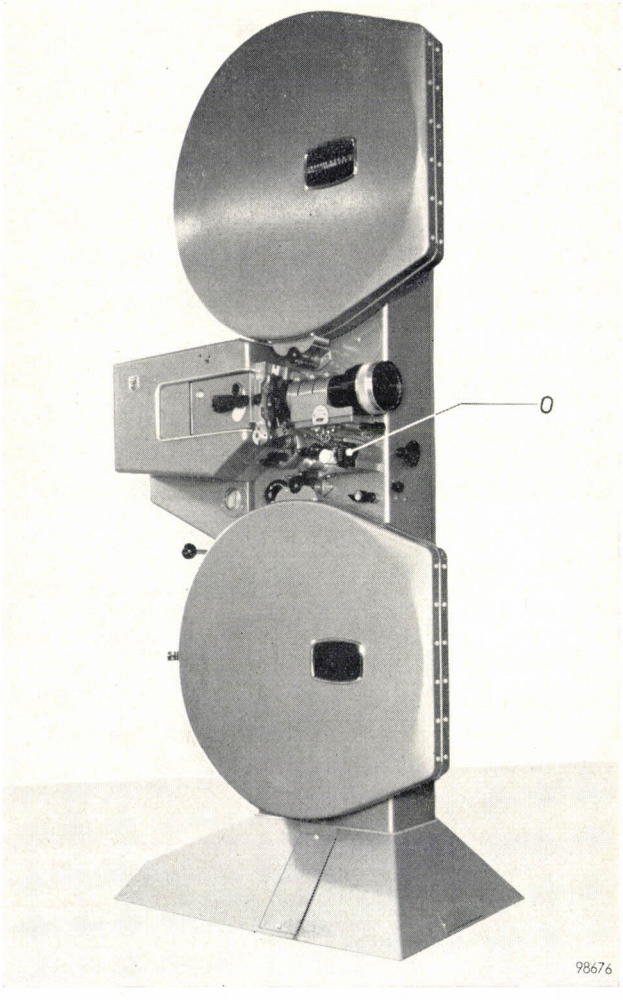
Principal design features

The conventional design

Present-day motion-picture projectors are almost invariably built-up from a number of sub-assemblies, each with its own function and as a rule fitting together on accurately machined horizontal faces. An example is the Philips FP 56 projector, shown in *fig. 1a*. This construction has the merit of enabling a projector to be adapted to technical advances by adding new sub-assemblies. When sound films came in, for instance, an optical sound-



a



b

Fig. 1. *a*) The Philips projector type FP 56 is an example of the conventional construction of horizontally fitting sub-assemblies. Mounted successively, one upon the other are: the foot *A*, the mounting platform *B*, the projection unit *C*, the magnetic sound-head *M*, and the upper spool-box *D*. Below the platform the lower spool-box *E* and behind the projection unit the lamp house *F* and the optical sound-head *O*.

b)The new Philips projector type FP 20. All components concerned with the film path are mounted on a flat vertical panel formed by the front face of a rectangular column. The projector is here equipped with a lamp house containing a pulsed light-source [2]) (it is then designated FP 20 S) and 6000′ spool-boxes. The optical sound-head is mounted at position *O*.

scanning system was added as a self-contained unit. With the advent of the magnetic sound track some years ago, a new sub-assembly was incorporated between the upper spool-box and the projection unit to make existing projectors meet the new requirements. The flexibility of the sub-assembly construction has thus allowed cinema proprietors to limit the costs of keeping abreast of technical innovations. It is evident, however, that adaptation along these lines does not lead to the most economical and functional design. Each sub-assembly needs its own cast-iron housing or frame, with horizontal faces that all require careful machining. Furthermore, since the path followed by the film must lie exactly in a single (vertical) plane, all these sub-assemblies must be painstakingly "lined up", which takes considerable time and trouble, even for experts [1]).

The new design

In the new projector (fig. 1*b*) a different approach was adopted. The main object was to simplify the design whilst retaining the advantage of flexibility, i.e. the possibility of adding or replacing components to meet the demands of the users. In short, the aim was to preserve the virtues of the sub-assembly layout but to avoid its drawbacks.

To this end, all components concerned with the film path are mounted on a *vertical* panel, thus dispensing with the laborious work of lining-up. The heart of the projector is a column of rectangular cross-section, bent and welded to shape from sheet steel. The column is closed at the top by a welded steel cover to increase the rigidity of the structure. At the bottom, which is open, a base piece gives it the required rigidity. The wide side-panel on which the components are mounted is easy to bring within the specified flatness tolerances. It requires no special machining and the assembly surfaces may even have a coating of paint.

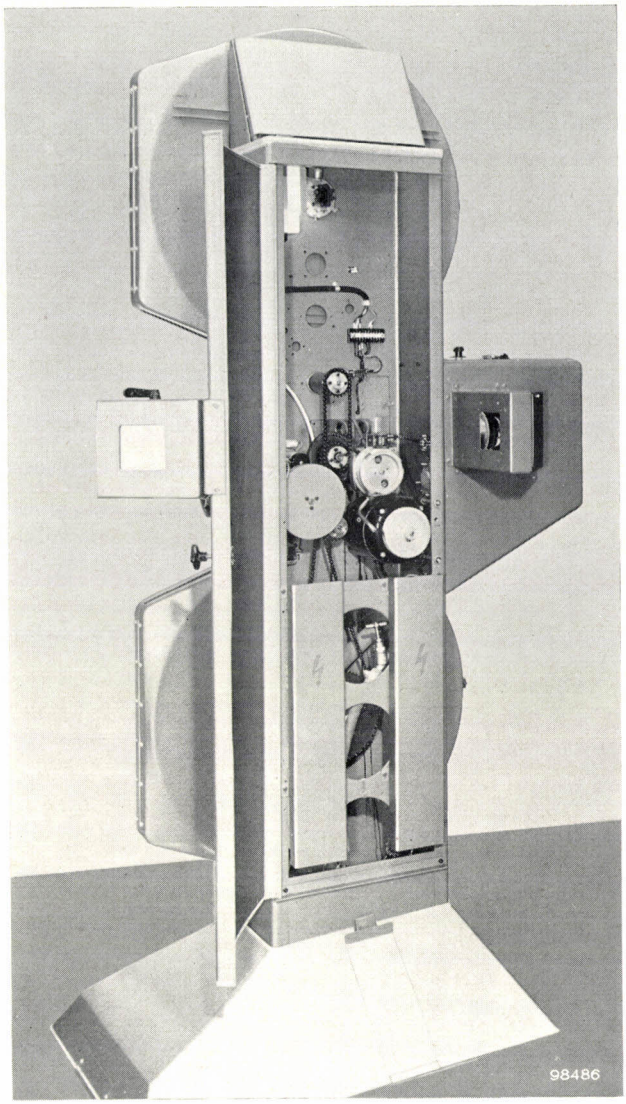


Fig. 3. The interior of the column is accessible through generous openings in the rear wall, closed by a common door. Note the chain-drive transmission system.

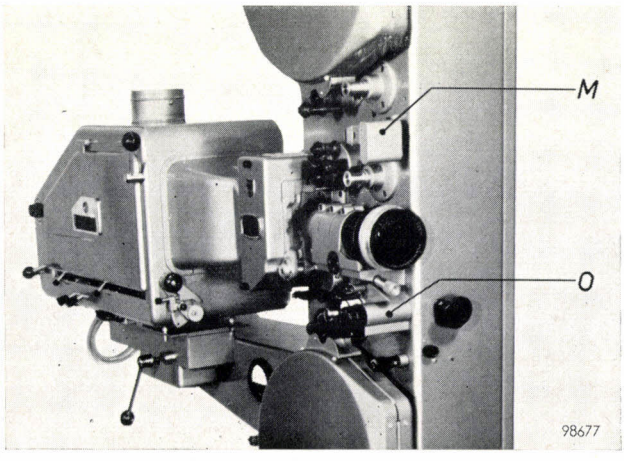


Fig. 2. The new projector equipped with an arc lamp (then designated FP 20), 2000′ spool-boxes and a magnetic sound-head *M* in addition to the optical sound-head *O*.

The mounting panel contains all the necessary holes, threaded and otherwise, appropriate to the various versions of the projector. For example, spool-boxes for 600 m (2000 ft) film spools can be mounted, or for 1800 m (6000 ft) spools. The bearing brackets for the spool spindles must then be shifted. In all versions the projector is fitted with an optical sound-head, but if the upper spool-box is moved up 140 mm there is room on the column for a magnetic sound-head too. The newly dev-

[1]) If a cinema has to be provided with a new projection system, the installation time is a point of a importance in that it determines the number of performances missed. The time needed to line-up the projectors adds significantly to the total time of installation.

eloped lamp-house with pulsed light-source, discussed in the preceding article [2]), can be attached directly on to the mounting panel. The projector shown in fig. 1*b* is equipped with the new lamp-house and 1800 m spool-boxes, but not with a magnetic sound-head. (The latter is, however, to be seen fitted on the projector in fig. 1 of the preceding article.)

Equipped with a pulsed light-source the projector is designated as type FP 20 S. If the light-source is an arc lamp — as in *fig. 2* — a platform is screwed to the column on which any current type of lamp-house can be mounted. The type designation is then FP 20. In fig. 2 the projector is fitted with a magnetic sound-head and 600 m spool-boxes.

The column construction makes it possible to drive all parts concerned by very simple means, that is by chains. The chains run slowly and silently, and are safely accomodated, free from dust, inside the column (*fig. 3*). In a projector as in fig. 1, on the other hand, two helical gears, two bevel gears and two transmission shafts are required for driving the lower spool alone. To change over from 600 m to 1800 m spools on the new projector, all that is necessary is to shift the bearing brackets for the lower spool spindle and to lengthen the driving chain. In the old construction it is necessary to mount, besides the bigger spool boxes, larger castings and a longer transmission shaft; the film path as well as the transmission shaft then has to be lined up.

The interior of the column is also the ideal housing for the electric wiring. The electrical connections between the various sub-assemblies by means of flexible cables (visible in fig. 1*a*) are thus a thing of the past. Since the column is open at the foot, the projector can be readily connected to the power supply and water main via lines under the floor.

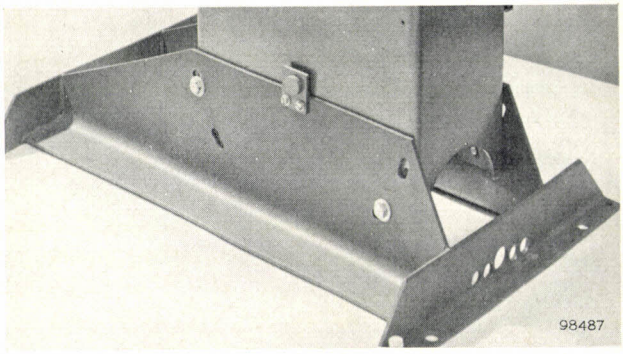


Fig. 4. The projector base with hoods removed.

The foot of the column fits into a base consisting of two angled plates welded to two L-sections to form a frame (*fig. 4*). The column rests on the plates at the protruding ends of a fixed spindle passing crosswise through the foot; the whole column can be tilted around this spindle into the required projection angle and then secured by bolts. The base frame is afterwards covered by a two-piece hood.

New design details

Framing adjustment

Since 35 mm film is provided at either side with four sprocket holes per frame, it is possible that a film may be positioned in the film gate one or two perforation holes too high or too low. It may also

[2]) See previous article in this issue: P. Hoekstra and C. Meyer, Motion-picture projection with a pulsed light-source, Philips tech. Rev. **21**, 73-82, 1959/60.

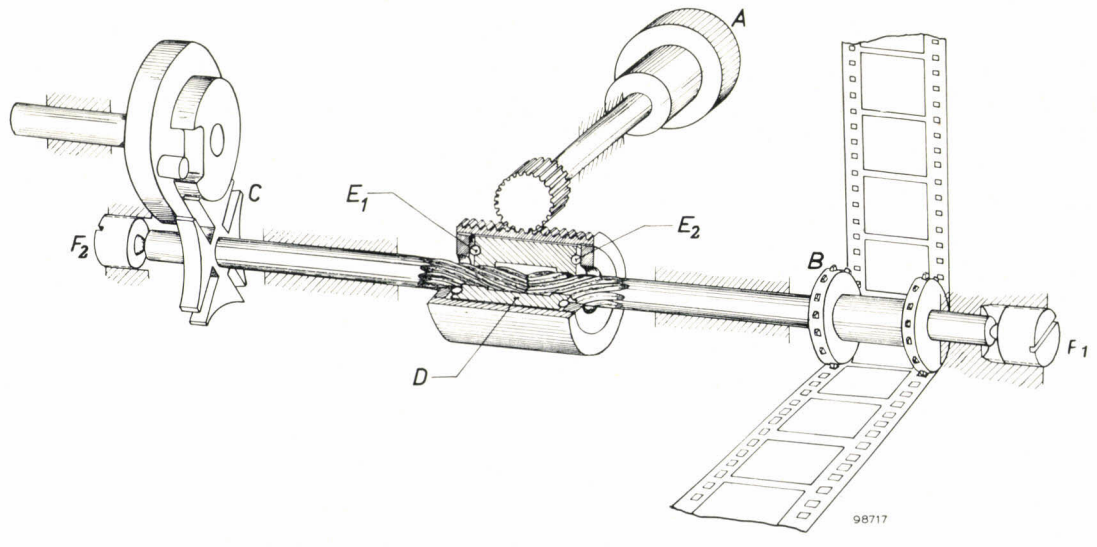


Fig. 5. Principle of the framing device used in the FP 20 projector. *A* framing knob, *B* intermittent sprocket, *C* Maltese cross, *D* coupling sleeve, E_1 and E_2 ball-bearings, F_1 and F_2 hardened thrust bolts.

happen during projection that when a badly framed splice has passed the film gate the black line dividing the frames appears on the projection screen. In order to correct such faults quickly, all 35 mm projectors are fitted with a device for framing adjustment.

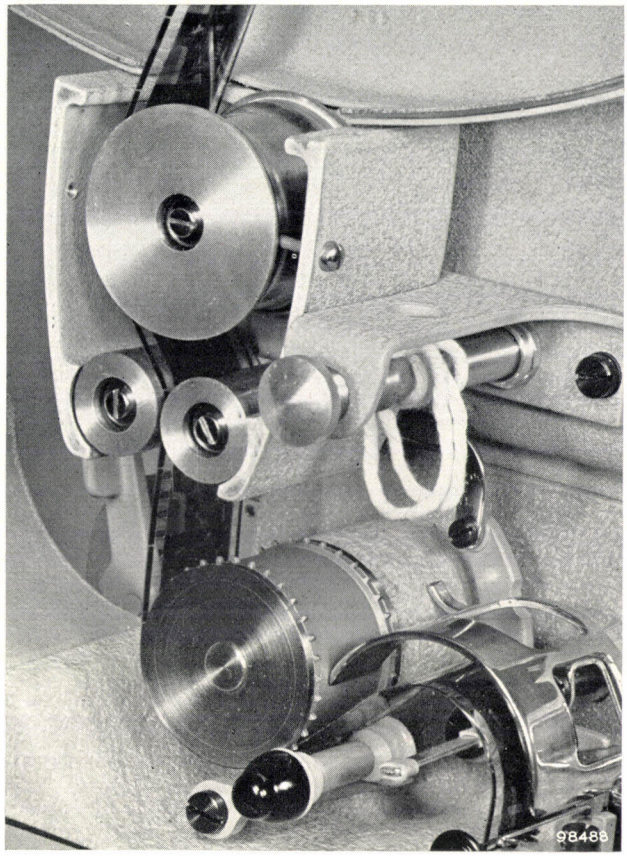
The framing device in the projector under discussion is of new design, the principle of which is illustrated in *fig. 5*. When the framing knob A is turned, the intermittent sprocket B undergoes rotation relative to the Maltese cross C, thereby shifting the film whilst in motion and effecting the necessary correction.

Many projectors in common use, including some Philips types, employ the method of central framing adjustment because of its reliability. Here, too, the intermittent sprocket turns during framing, but since the spindle between the Maltese cross and intermittent sprocket is not divided, the whole Maltese cross mechanism turns too. If no counter-measures are taken, this upsets the mutual adjustment between the frame-shift period and the light-interception by the shutter. A complicated mechanism is required to prevent this happening.

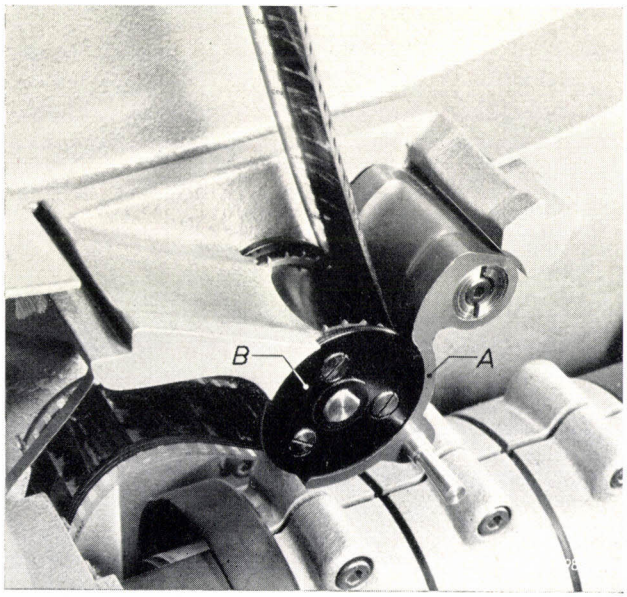
The coupling sleeve D (fig. 5), which is internally splined to mate the coupled ends of the spindle, is an injection-moulded product of nylon. This is an ideal material for the purpose in question, where a sleeve is required that can be shifted along a shaft with absolutely no backlash [3]). The dimensions of the sleeve can quite permissibly be made 1% to 2% smaller than the spindle (with steel-to-steel, a press-fit is obtained with under-dimensioning of only 0.01-0.1%. Initially a good deal of force is needed to shift the sleeve, but in the course of about 24 hours the material "sets" as a result of creep, so that the force required drops to an acceptable level. This force is transmitted to the sleeve via thrust ball-bearings E_1 and E_2 and finally taken up by the hardened thrust bearings F_1 and F_2 at both ends of the divided intermittent spindle. Life tests have shown that no backlash occurs even after intensive use.

Fire traps

The fire traps are another example of simplified design in the new projector. Their purpose is to ensure that, if the film catches fire, the fire will not spread to the spools. The conventional design of a fire trap (at the upper spool-box) can be seen in *fig. 6a*. Upon leaving the feed spool the film passes through a roller fire-trap which is closed at the front by a lid fastened to the spool box (not visible in the photograph). Below the fire trap is a sprocket with two pad-rollers which ensure that the sprocket teeth engage properly in the film perforations. Fig. 6*b* shows the new design in which the pad rollers are replaced by a guide shoe enveloping the sprocket



a



b

Fig. 6. *a*) Illustration of conventional fire-trap design as employed in projector type FP 56.
b) Fire-trap design in the FP 20 projector, composed of guide shoe and sprocket.

[3]) The properties of nylon as an engineering material are discussed e.g. in A. J. Cheney, Designing with nylon, Machine Design, 23rd February 1956, 95-102, and 8th March 1956, 95-99.

over a wide angle, such that shoe and sprocket together constitute at the same time a fire trap. Here too the fire trap is closed by a flat lid attached to the spool cover (not visible in the figure). This design cuts down on components, dispenses with the need for alignment and halves the time taken to thread the film.

Remote control from the auditorium

In large cinemas the screen is so far away from the projectors that the projectionist is unable to focus the picture as sharply as someone nearer the screen. Provision has therefore been made in the new projector for incorporating inside the column a small motor which, via reduction gears, controls the focus adjustment. The motor is operated from a control box set up inside the auditorium. The framing and the sound volume can also be adjusted from this control box. If the projector is equipped with the new pulsed-source lighting system, the box can also be provided with a control for regulating the lamp power [4]).

[4]) See preceding article, page 80.

Summary. Conventional motion-picture projectors are composed of self-contained sub-assemblies which fit together on accurately-machined horizontal faces. A new projector of simpler design is described. All components concerned with the film path are mounted on the flat front panel of a sheet-steel column of rectangular cross-section. This dispenses with the laborious work of lining-up the film path when installing the projector. The front panel contains all holes, threaded and otherwise, required for adapting the projector with the minimum of trouble to individual requirements. The column accomodates a simple and reliable chain-drive transmission system; it is also the ideal housing for the electric wiring. Some other design details are discussed, viz. a framing device and a fire-trap system, both of simplified design.

A SLOTTED LECHER LINE FOR IMPEDANCE MEASUREMENTS IN THE METRIC AND DECIMETRIC WAVE BANDS

by G. SCHIEFER*)

621.317.332.1:621.372.2

In the V.H.F. bands, impedance measurements on balanced components such as aerials, transformers and coupling loops, are usually made with unbalanced (unsymmetrical) test equipment, special fourpoles or "baluns" being inserted to effect the transition from balance to unbalance. Generally speaking such inserted devices are open to the following objections: either they have a very narrow frequency band and must accordingly be matched with the utmost accuracy to each test frequency (e.g. half-wave stubs), or they have a broad frequency band but at the same time such complex four-pole properties that impedance measurements are only possible if the precision required is not very high. For accurately measuring balanced impedances (in particular low-loss reactances) within a wide range of frequencies we have therefore designed a slotted lecher line on which the standing wave pattern can be detected in the familiar way with a travelling probe and thus the connected impedance derived.

The design of such a slotted line is governed primarily by the following considerations:

a) The line must be screened to minimize interference due to "hand effect" and radiation pick-up (e.g. in aerial measurements).

b) The shape of the cross-section must be such that the characteristic impedance of the line may be calculated beforehand from the geometry.

c) The characteristic impedance must be affected as little as possible by slight errors in the position of the conductors.

d) It should be possible to lengthen the line at the object end with ordinary lecher wires, whether screened or not, without causing reflections at the junction as a result of cross-sectional disparities.

e) The useful length of the line must be greater than half the longest-occurring wavelength.

Conditions *a*) to *d*) point in the direction of a screened lecher line with inner conductors and screening of circular cross-section (*fig. 1*). The characteristic impedance of such a line may be calculated very accurately, and at a given ratio d/D of the diameters it shows a maximum when the centre-to-centre distance a of the inner conductors is approximately half the inside diameter D of the screen. In that case the electrical properties of the line are insensitive to small errors in the position of the inner conductors. Since the characteristic impedance did not have to have any specified value, $d = \frac{1}{4}D$ was decided upon. This leaves a wide choice for d and D amongst the brass tubing commercially available. Moreover the damping of a screened lecher line is just about minimum with this ratio of diameters. Calculated according to Sommer [1]), the characteristic impedance proves to be 104.7 ohms; allowing for the narrow lengthwise slot (see below) the value is 105 ohms. (A round figure of e.g. 100 ohms would have simplified measurements, but would have led to an unfavourable value of d/D.)

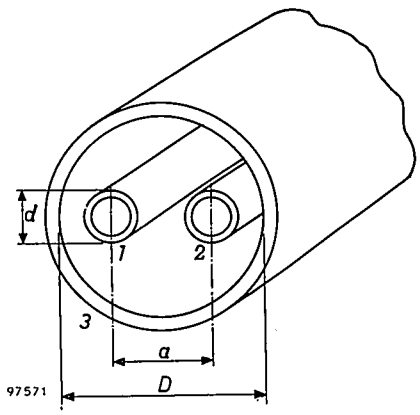


Fig. 1. Lecher line *1-2* with outer conductor (screen) *3*. Relations between dimensions: $a = \frac{1}{2}D$, $d = \frac{1}{4}D$.

In order to carry out measurements in the whole of V.H.F. band II (87.5-100 Mc/s, maximum wavelength 3.45 m), the line is made about 2 metres long; see point *e*) above. This makes the lowest measuring frequency about 80 Mc/s. Of course, measurements may also be extended to lower frequencies by connecting additional sections of line. The upper frequency limit of the line is set by the occurrence of waveguide modes of oscillation which, however, appear only at frequencies above 1000 Mc/s.

The line (*fig. 2*) consists of a rigid brass tube (inside diameter 48 mm, outside diameter 60 mm) which serves as the screening conductor, and two inner conductors, also of brass tubing (10 mm inside and 12 mm outside diameter). The inner conductors are held by supports mounted at distances of about

*) Zentrallaboratorium Allgemeine Deutsche Philips Industrie GmbH, Aachen laboratory.

[1]) F. Sommer, Die Berechnung der Kapazitäten bei Kabeln mit einfachem Querschnitt, Elektr. Nachr.-Techn. **17**, 281-294, 1940.

40 cm apart; each support consists of two 4 mm bolts of "Teflon" (polytetrafluorethylene), which are screwed at right-angles to each other into the inner conductors, their heads being locked in the outside conductor. At the line input the support is a disc of "Trolitul", which is permissible since reflections are not critical at this location. The screen also serves as a guide rail for the carriage carrying the probe-detector; the latter, which we shall discuss presently, projects into the line through a 5 mm-wide slot cut into the screen along the whole of its length.

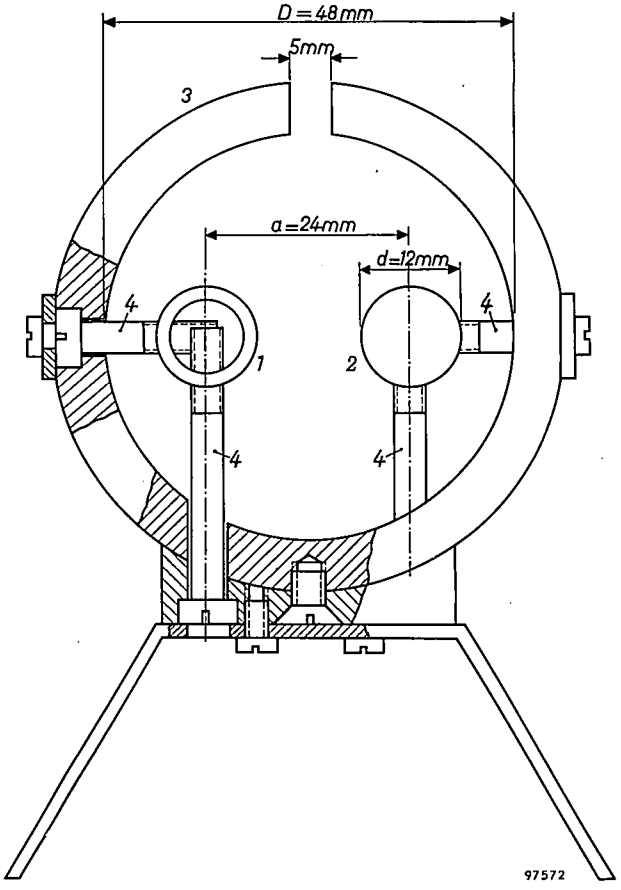


Fig. 2. Cross-sectional view of the lecher line. *1* and *2* conductors of brass tubing, supported by bolts *4* of "Teflon". *3* screen of brass tubing with slot for probe.

As fig. 2 shows, this design makes it easy to adjust the horizontal and vertical alignment of the inner conductors; half the space inside the screen in which the probe moves is completely free of interfering elements. The effect of the supports on the electrical properties of the line is thereby negligible. This is due in large measure to the low dielectric constant of the plastic insulation material used ($\varepsilon_r = 2$). The mechanical properties of the material have also proved satisfactory; after six months' use there was no sign whatsoever of creep deformation.

The line is fitted at both ends with symmetrical connection flanges (see fig. 4) with plug-sockets for the inner conductors.

Special attention was paid to the probe-detector. With all balanced line systems it is necessary in practice to take into account the occurrence of asymmetrical waves. These arise even if there is only a slight unbalance or asymmetry somewhere in the transmission system, and if resonance conditions are favourable they may be exceptionally severe. In the case of a screened lecher line the danger is particularly great, for it also possess good transmission properties as a *coaxial* system, the inner conductors then being in phase with each other and the screen in antiphase.

Where asymmetrical waves arise as a result of unbalance in the test object itself, they inevitably result in spurious measurements. It is easy to see that the balanced impedance of a two-terminal network can only be measured properly by means of a three-conductor system if the third conductor (the outer one) remains neutral. Such errors of measurement can therefore be avoided only by ensuring that the test objects are accurately balanced electrically, which in most cases also means geometrically symmetrical.

Where, however, asymmetrical waves arise as a result of unbalance preceding the test object — i.e. mainly in consequence of an unbalanced supply voltage — they can be prevented from affecting the result of the measurement if a probe be used that is insensitive to these waves. With this in mind we first made a series of experiments with tuned probes, which seemed to us favourable because of their high sensitivity and their suppression of higher harmonics. We found, however, that neither with electrical nor with magnetic coupling was it possible, at reasonable cost, to make the tuning device sufficiently balanced to suppress the indication of unbalanced waves in the whole frequency band. In this respect non-tuned probes were better, but they were not sensitive enough. Satisfactory results were finally obtained by introducing the detector diode of the probe directly into the radio-frequency field of the two inner conductors. In this way we dispense with all connections in the R.F. circuit of the probe that might give rise to frequency-dependence and unbalance. At the same time we have a probe that is sensitive enough and relatively easy to balance in the whole frequency band.

The final arrangement is shown schematically in *fig. 3*. The connection wires of a miniature germanium diode, type OA95, form a dipole located directly in the field between the two inner conductors.

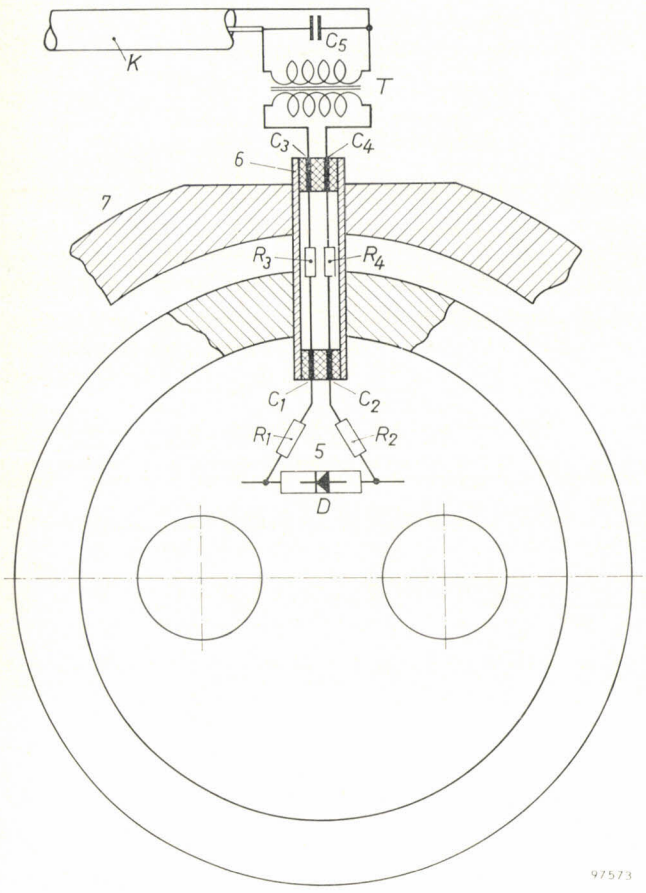


Fig. 3. Probe arrangement. The probe proper *5* is fixed by means of tube *6* to the carriage *7* which travels along the line. *D* miniature germanium diode OA95. R_1-R_2 miniature resistors constituting the high-frequency load for the diode. The balanced lead-through capacitors C_1-C_2 and C_3-C_4, together with resistors R_3-R_4, form a low-pass filter. *T* transformer tuned to 1000 c/s. C_5 matching capacitor at low-frequency side. *K* cable to probe amplifier.

The current is taken off via two symmetrically connected miniature resistors, which also act as the R.F. load for the diode. The high frequency is filtered-out by a balanced low-pass filter, consisting of special lead-through capacitors and resistors that form an intrinsic part of the tube by which the probe is introduced through the slot.

To make the indication as sensitive as possible the high-frequency supply voltage is modulated in amplitude by 1000 c/s, and a transformer tuned to 1000 c/s is incorporated in the low-frequency part of the probe for the purpose of matching to the unbalanced input of the probe amplifier. At a bandwidth of 8 c/s this specially designed amplifier is so sensitive that a low-frequency voltage of 0.5 μF can still be read-off with certainty.

In this way the total sensitivity is such that, at a high-frequency supply voltage of 5 V (across 105 ohms) and a standing-wave ratio of some hundreds, the voltage minima can still be readily detected. This is just about the limit that can be reached, seeing that standing-wave ratios of this order of magnitude are caused by the natural damping of the line itself (when short-circuited). However, since the probe has a wide frequency-band, such measurements call for a high-frequency supply voltage substantially free from higher harmonics.

The probe is mechanically guided by a carriage, as mentioned, which travels smoothly along the screen on six ball bearings. Also mounted on the carriage are a device for reading the length coordinate and a box containing the matching trans-

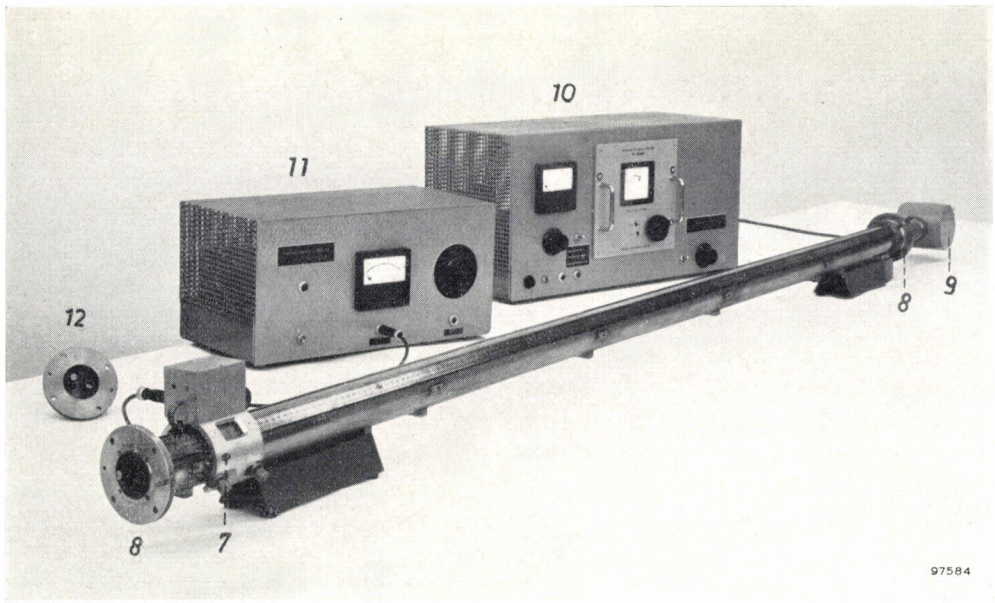


Fig. 4. Complete equipment. In the foreground the transmission line with connection flanges *8* and balun *9*. Left, on the line, the probe carriage *7* travelling on ball bearings. At the back the generator *10*, the probe amplifier *11* and a matched termination impedance *12* of 105 ohms.

former, with a socket for connecting the cable to the probe amplifier. After careful alignment of the inner conductors and proper positioning of the probe, variations in the indication as a result of errors in the parallel travel of the carriage with respect to the inner conductors are less than 2%.

Fig. 4 shows the line, complete with the high-frequency generator and the probe amplifier. The principal data are given below. The apparatus is at present used in the frequency range from 80 to 300 Mc/s.

Principal data of the balanced slotted line

Line

Characteristic impedance	105 ohms ± 0.5%.
Natural damping	at 100 Mc/s: 9×10^{-6} Np/cm at 300 Mc/s: 16×10^{-6} Np/cm
Useful length:	1891 mm.

Probe

Variation in sensitivity between 80 and 300 Mc/s	< 2 : 1.
Disturbance introduced by probe in the R.F. field, in maximum of standing wave	< 1%.
Diode characteristic:	quadratic to a low-frequency indication of approx. 1 mV.

Longitudinal uniformity: mechanical tolerances of the line and probe guide system, and inhomogeneities such as supports and connections, give rise to non-uniformities along the length of the line. Their effect on the measurement of voltage and length is determined as follows. With a matched termination, consisting of a balanced resistance of 105 ohms, the indicated voltage over the total length of the line varies by a maximum of 2%. If the line is terminated by a shorting plunger, the voltage nodes shift by ± 1 mm max. The absolute accuracy in impedance measurements can be derived from these data from case to case.

High-frequency generator

Frequency range	80-300 Mc/s.
Power output	1 W in unbalanced resistance of 60 ohms.
Higher harmonics.	< 1%.
Modulation frequency	1000 c/s ± 1 c/s.
Modulation depth.	> 80%.

The high-frequency energy is conducted to the test line via a broad-band balun with conical transition.

Probe amplifier

Frequency band	1000 c/s ± 4 c/s.
Input impedance	100 kΩ.
Measuring ranges	12 ranges with full scale deflections of 3 μV to 1 V.
Noise voltage	< 0.15 μV.

Accessories

For connecting test objects to the line, extension pieces are available and also adaptors for changing over to smaller cross-sections. Careful construction and correction of the supports ensure that these accessories cause no measurable mismatch or unbalance errors.

Summary. For impedance measurements on balanced objects in the V.H.F. bands (80-300 Mc/s), a balanced, screened transmission line about 2 metres long has been designed in the Philips laboratory at Aachen. The characteristic impedance is approx. 105 ohms. The probe is insensitive to unsymmetrical waves, the detector diode (a miniature germanium diode OA95) being introduced directly into the R.F. field inside the line. The high-frequency supply voltage is modulated in amplitude at 1000 c/s. The total sensitivity in such that, at a high-frequency supply voltage of 5 V and a standing wave ratio of some hundreds, the voltage minima can still be accurately measured.

AN 8 mm HIGH-RESOLUTION RADAR INSTALLATION

by J. M. G. SEPPEN *) and J. VERSTRATEN **).

621.396.967.029.65

The display on a radar screen of the PPI type (plan position indicator) represents an "echo map" of the area scanned by the aerial. With relatively long waves, objects can be "seen" in this way which are far beyond the optical range. Radar is also used, however (with shorter waves), to detect and locate objects within the optical range, for darkness and fog present no obstacles to radio waves.

An area viewed by direct optical means obviously contains many more details than the radar display, which is merely a very rough indication. The degree to which objects can be distinguished, i.e. the resolution of the apparatus, sets a limit to the usefulness of radar for various purposes. N.V. Philips Telecommunicatie Industrie have now designed a radar installation, type 8 GR 250, which operates in the 8 mm wave range and has a very high resolution. The article below gives a description of this apparatus together with some theoretical considerations underlying the design.

Introduction

The 3 cm and 10 cm wavelengths have long been used for navigational radar with excellent results where fairly long ranges are concerned. For navigation within a restricted area, however, the apparatus is required to have a very high resolution, which is difficult to achieve at these wavelengths. In harbours and narrow waters, for example, navigators not only require accurate knowledge of the distances to the numerous obstacles, they also wish to see the manoeuvres being carried out by other ships. The latter can be deduced if the true shape of the ships is discernible on the radar screen ("ship-shape" radar).

A resolution as high as this is only possible with a radar set built to operate at an even shorter wavelength. N.V. Philips Telecommunicatie Industrie have designed such an installation, type 8 GR 250, working on a wavelength of 8.6 mm. Apart from its usefulness on ships frequently in narrow or busy waters, or on large ships as auxiliary and harbour radar, it can also be used in airports for controlling traffic on runways and aprons (ASMI-radar, Airfield Surface Movement Indicator [1]).

Principles of radar

Since the subject of radar [2]) has never been dealt with at any length in this journal, we shall begin with a brief description of its principles.

The radar station sends out radio-frequency pulses which, being reflected from the objects in their path, give rise to echo signals which are picked up by the radar receiving system and displayed (for example) on the screen of a PPI tube (Plan Position Indicator). The echoes then trace out on the screen a "map" of the area scanned, the objects appearing on the screen as luminous spots on a dark background. We shall now examine in greater detail how such a system works, with reference to the block diagram of the 8 mm radar installation given in *fig. 1*.

Periodic high-voltage pulses from a modulator are applied to an 8 mm magnetron which thereupon delivers high-frequency signals of extremely short duration. The repetition frequency is controlled by a pulse generator which supplies the synchronizing signals. The high-frequency pulses delivered by the magnetron (wave trains) are conducted via a waveguide to the antenna. This consists of a parabolic reflector at the focal point of which the waveguide is flared out to form the "feeder" (horn). The wave trains are transmitted by the reflector in a narrow beam. If they encounter an obstacle in their path, energy is reflected (i.e. scattered) in all directions. The signals reflected back to the radar station are picked up by the receiving antenna, which is mounted parallel with the transmitting antenna. A superheterodyne receiver, containing a mixer stage and a klystron as local oscillator, converts the echo signals into pulses of lower frequency (intermediate frequency, in this case 90 Mc/s). These are amplified, detected, again amplified in a video amplifier and then applied to the cathode-ray tube. A synchronizing pulse obtained from the magnetron ensures that the time base of the CRT starts at the same

*) N.V. Philips' Telecommunicatie-Industrie, Hilversum.
**) Formerly with N.V. Philips' Telecommunicatie-Industrie.

1) This 8 mm radar equipment has already been described by the authors in T. Ned. Radiogenootschap **23**, 17-32, 1958 (No. 1) and in Philips Telecomm. Rev. **20**, 5-15, 1958, (No. 1).

2) Coined from "*r*adio *d*etection *a*nd *r*anging". See e.g. Louis N. Ridenour, Radar system engineering, Radiation Laboratory Series, No. 1, McGraw-Hill, New York 1947.

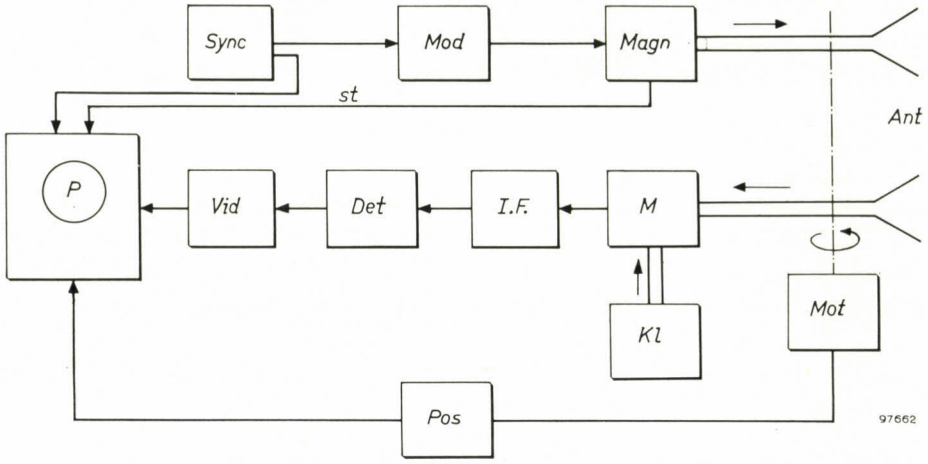


Figure 1. Block diagram of the 8 mm radar installation. *Magn* = magnetron, *Mod* = modulator, *Sync* = synchronizing pulse generator, *Ant* = rotating antennae, *Kl* = klystron, *M* = mixer, *I.F.* = intermediate frequency amplifier, *Det* = detector, *Vid* = video amplifier, *P* = PPI unit, *Mot* = motor for driving the antennae, *Pos* = position-coupling of antennae and time base; *st* denotes the starting pulse delivered by the magnetron to the time base circuit.

moment as the transmitted pulse. The electron beam is magnetically deflected in the radial direction by a coil made to rotate around the neck of the tube in synchronism with the antennae, which rotate about a vertical axis. The time base of the CRT begins in the middle of the PPI screen and runs radially to the edge in a direction corresponding to that in which the antennae are pointing. The received pulse is now used to modulate the intensity of the electron beam, which is normally just below the threshold of illumination. In this way a map of the swept area is traced out on the PPI, echoes from objects farther away taking proportionately longer to reach the receiver and thus producing a luminous spot farther away from the centre of the PPI screen. *Figs. 2* and *3* give some idea of the accuracy with which the PPI of the 8 mm radar installation displays the surroundings.

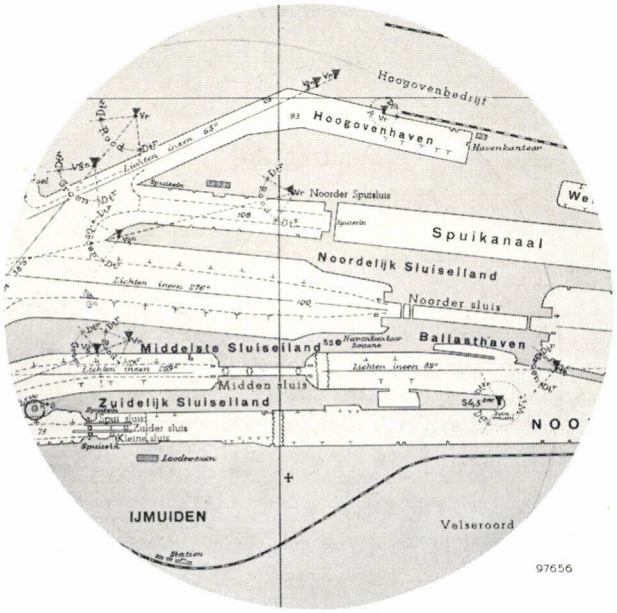


Fig. 2

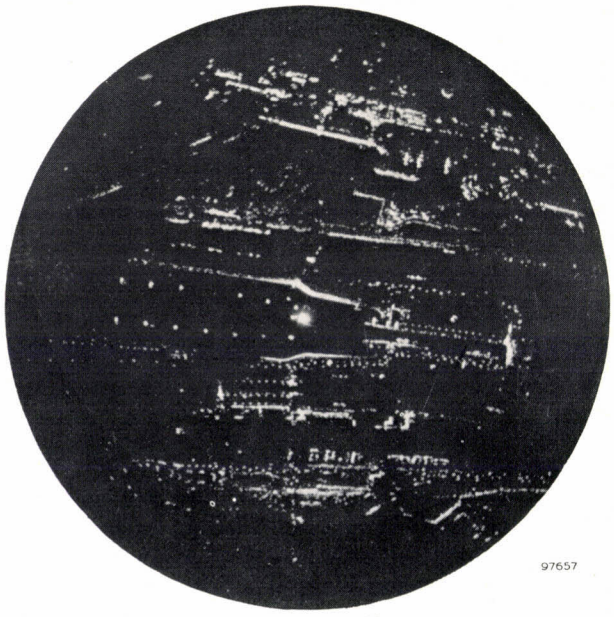


Fig. 3

Fig. 2. Map of Ymuiden docks, scale 1 : 25 000.

Fig. 3. Radar display of Ymuiden docks obtained with the 8 mm radar installation 8 GR 250. (Screen coverage set to 2 km diameter.)

The operation of radar thus depends among other things on the following three facts [3]):

1) The velocity of propagation of electromagnetic waves is constant, so that the time taken for a pulse to travel to an object and back is a direct measure of the distance to the object.
2) The transmitter is able to generate pulses of high power and extremely short duration. In the interval between these pulses the receiver has the opportunity to register the echoes.
3) The use of very short wavelengths (in the centimetre or millimetre range) makes it possible to produce a very narrow beam of radiation with an antenna of moderate size, so that the pulses are transmitted and received in a well-defined direction.

Effect of wavelength on the properties of a radar system

The foregoing indicates that the properties of a radar system are largely determined by the wavelength. In the following we shall discuss the effect of wavelength on the resolution (discrimination) of the system and on the range.

Antenna

As we have seen, the antenna is a parabolic reflector fed with RF energy via a horn at the focal point. A sharp parallel beam is not produced; owing to diffraction phenomena it fans out into a more or less conical pencil with its origin at the centre-point of the antenna. At the edge of this pencil the field-strength falls off fairly rapidly to zero.

The apex angle of the pencil in the horizontal plane is conventionally defined as the angular width ϑ_h between the directions in the beam in which the intensity of the radiation falls to half the maximum intensity. This beam width (in radians) is approximately given by the formula:

$$\vartheta_h = \lambda/l_h, \quad \ldots\ldots \quad (1a)$$

where λ is the wavelength and l_h the linear dimension of the antenna in the horizontal plane. In the same way one can also define a beam width in the vertical plane (see *fig. 4*):

$$\vartheta_v = \lambda/l_v. \quad \ldots\ldots \quad (1b)$$

For a measure of the directivity of an antenna, the radiation in a certain direction is compared with that of a hypothetical isotropic source, i.e. one which transmits energy equally in all directions; if this source radiates a total power P_0, a power $P_0/4\pi R^2$ will pass through unit area at a distance R.

[3]) We are concerned here solely with pulse radar. Other systems exist, but are not relevant in this connection.

With the aid of a reflector the intensity in a given direction can be increased by a factor G. The antenna power gain or directive gain is taken as the maximum value of this factor, that is, along the axis of the beam. It is given by the formula:

$$G_0 = 4\pi A/\lambda^2, \quad \ldots\ldots \quad (2)$$

where A is the effective area of the antenna (approximately 0.7 times the actual area). The way in which this formula is arrived at can be understood by regarding the solid angle as the product of the beam widths ϑ_h and ϑ_v (see equations $1a,b$) and by comparing this solid angle with the solid angle 4π in which the isotropic source radiates; hence $G_0 = 4\pi/\vartheta_h\vartheta_v$. From (2) it can be seen that a shorter wavelength results in an increased gain. The effect of this on the range of the radar installation appears in the radar equation discussed below.

The radar range equation

If the power radiated is P_0, the signal power received back at the radar station is given by

$$P = \left(G_0 \frac{P_0}{4\pi R^2}\right)\left(\frac{\sigma}{4\pi R^2}\right)A. \quad \ldots \quad (3)$$

This equation assumes propagation in free space, no account being taken of losses due to atmospheric effects, etc. The first factor comprises the above-mentioned transmitted power beamed by the antenna and incident on an object at a distance R. The

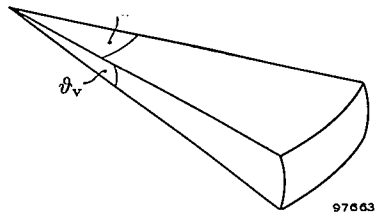


Fig. 4. Illustrating the horizontal and vertical beam widths

object scatters the incident energy, and a small fraction returns in the direction of the radar station. This fraction is expressed in the second factor, where σ is the effective reflecting area of the object.

Taking for P the value P_{min}, which is the minimum signal detectable by the receiver, we can write for the range R_{max} of the radar system:

$$\left.\begin{aligned} R_{max} &= H \sqrt[4]{\frac{P_0}{P_{min}}} \sqrt[4]{\frac{\sigma}{4\pi}}, \\ \text{where} \quad H &= \sqrt{\frac{G\lambda}{4\pi}} = \sqrt{\frac{\lambda}{\vartheta_h\vartheta_v}} = \sqrt{\frac{A}{\lambda}}. \end{aligned}\right\} \quad \ldots \quad (4)$$

From this "radar equation" we can draw the following conclusions.

a) Assuming the antenna dimensions, peak pulse power and minimum detectable signal to be kept constant, the range in free space is inversely proportional to $\sqrt{\lambda}$, i.e. proportional to $\sqrt{f}$, the root of the frequency. This would argue in favour of making the frequency as high as possible, were it not that electromagnetic-wave propagation in the atmosphere becomes, in general, poorer with increasing frequency. We shall return to this presently.

b) If the directive gain, and hence the beam width, of the antenna be kept constant — which implies that the linear dimensions of the antenna must change proportionately with the wavelength — then the range is proportional to $\sqrt{\lambda}$.

c) If the range is to be kept constant, assuming the properties of transmitter and receiver to be constant, then A/λ must remain constant; in other words, the linear dimensions of the antenna must be proportional to $\sqrt{\lambda}$.

Applying the above considerations to the transition from a 3 cm to an 8 mm radar system (and assuming for convenience that the ratio of the wavelengths is 4), we see the following. The beam width being the same, the range is reduced by half. The antenna area becomes 16 times smaller. Compensation of the loss of range by increasing the power will entail at least a 16 times higher power. If we wish to keep the range constant, the antenna of the 8 mm system must be made only 4 times smaller in area, i.e. half the linear size. It is then still necessary to compensate for a loss of range due to atmospheric effects.

Resolution

A distinction is made between tangential and radial resolution. On the PPI and over the area around the radar station we imagine there to be a system of polar coordinates with the origin at the radar station. The *radial* resolution is determined by the length τ of the pulse sent out by the magnetron. The transit time of the pulse from the antenna to an object at a distance R and back is given by $t = 2R/c$, where c is the velocity of light. Two objects, separated radially from each other by a distance ΔR, can only be distinguished if the difference in the transit time of the pulses, $2\Delta R/c$, is greater than the pulse length τ. This minimum distance, which is equal to $c\tau/2$, is the radial resolution, or discrimination, of the radar system. It is evidently desirable to have the shortest possible pulse length τ. This cannot be indefinitely shortened, however, since the magnetron requires a number of cycles before it gets up to peak power. This implies that at shorter wavelengths pulses of shorter duration can be generated, thus improving the radial resolution. With the 8 GR 250 the pulse length τ is 0.02 μsec, which corresponds to a radial resolution of 3 metres. (We disregard here various contingent factors affecting the resolution.)

The *tangential* resolution is given by the product of the horizontal beam-width ϑ_h and the distance R to the antenna: $R\vartheta_h = R\lambda/l_h$. In order to improve tangential resolution we can therefore either make the antenna larger or the wavelength shorter. With the 8 GR 250 the beam width is approximately 0.3°, and the tangential resolution at a distance of 600 metres approximately 3 metres.

A limitation set to the resolution of a radar system is the finite size of the luminous spot on the cathode-ray tube screen. On the PPI a "map" of the scanned area is reproduced. It is useless if the objects to be discriminated by radar are such that their screen image would be smaller than the luminous spot produced on the screen by the electron beam. In practice the image on the PPI is displayed at a scale adjustable in a number of steps. The largest scale corresponds to the smallest area displayed which, in the 8 GR 250, is a circle having a radius of about 500 metres. On the next scale the radius is about 1000 metres, and so on. The limiting effect of spot size is obviously least for the smallest area. With the 8 GR 250 the various parameters have been so selected as to achieve in the smallest areas the best possible compromise between the resolution of the radar system and that of the cathode-ray tube.

Pulse repetition frequency

The choice of the pulse repetition frequency F is also dependent on the wavelength used. This frequency has to be chosen between certain limiting values.

The maximum value is determined by the following factors.

a) The range R_{max} of the radar system. Considering that each transmitted pulse must be able to travel the maximum distance (to the target and back) before the next pulse is transmitted, the maximum repetition frequency is given by

$$F_{max} = \frac{1}{\frac{4}{3} t_{max}} = \frac{3}{4} \frac{c}{2R_{max}}. \quad . \; . \; . \; (5)$$

The correction factor $\frac{3}{4}$ allows approximately for the flyback time of the cathode-ray tube sweep.

b) The maximum average power W_0 delivered by the magnetron. This imposes the condition:

$$F_{max}\,\tau P_0 = W_0. \quad \ldots\ldots \quad (6)$$

The minimum repetition frequency is governed by the speed at which the antenna rotates and by the horizontal beam-width ϑ_h. Since there must be at least one transmitted pulse per angular displacement ϑ_h of the antenna, it follows that

$$F_{min} = \frac{2\pi N}{60\,\vartheta_h}, \quad \ldots\ldots \quad (7)$$

where N is the number of antenna revolutions per minute. In order to be able to distinguish weak echoes on the screen from the random fluctuations caused by receiver noise, the pulse repetition frequency is always at least several times higher than F_{min}. The antenna then scans each perceptible object with more than one pulse at a time.

With the aid of formulae (4) and (1) we can conclude from the above that, where the antenna dimensions are constant, F_{max} is proportional to $\sqrt{\lambda}$, and F_{min} inversely proportional to λ. A shorter wavelength thus means a lower F_{max} and a higher F_{min}. In the choice of F the value of F_{min} is of more importance in as much as the number of pulses per observed object has a very marked influence on the minimum detectable signal P_{min}.

Since the 8 GR 250 equipment is especially intended for short-range use, e.g. for surface movement indication on airfields and for navigation in harbours and narrow waterways, a high speed of antenna revolution is evidently desirable, for at distances of a few hundred metres the angular velocities at which objects can change their position with respect to the radar post can be quite considerable. For this reason, and also because of the short wavelength used, F_{min} had to be fairly high. This led to the choice of $F = 5000$ c/s.

Information capacity

In the above we have indicated how the properties of a radar system depend on the wavelength used. We have seen, among other things, that the resolution and theoretically the range too (that is, neglecting absorption) of the system increase with increasing frequency. We shall now consider the rate at which radar information can be collected, i.e. the information capacity. This is a more general indication of the performance of the system than range and resolution. With the aid of the foregoing considerations we shall arrive at a general formula with which we can ascertain the effect which various quantities, in particular the frequency, have on the information capacity of the system.

We shall first determine the number of observable objects (see *fig. 5*).

No matter how small an object is, it cannot be reproduced smaller than $c\tau/2$ in the radial direction or smaller than ϑ_h in the tangential direction. The maximum number of objects observable within a radius equal to the range R_{max} (defined for objects with a given reflecting area σ) is therefore:

$$n = \frac{2\pi}{\vartheta_h}\,\frac{R_{max}}{c\tau/2}. \quad \ldots\ldots\ldots \quad (8)$$

According to information theory we should next ascertain the number of observable states to which one object can give rise. Depending upon its distance, size, etc., an object can return signals to the receiver that vary enormously in strength. In practice, however, this variation is virtually not perceptible in the display (particularly not on the PPI). We can therefore simply say that the number of perceptible states of an object is equal to two, i.e. present or absent — or, more exactly, the state either does or does not give rise to a reflected power at the receiver in excess of the minimum detectable power. With two observable states for each of n objects the total quantity of information I is equal to n bits.

For each complete revolution of the aerial the radar system can thus deal with an amount of information I as given by (8). The information capacity C is the amount of information that can be handled per unit time. If N be the number of antenna revolutions per minute, then

$$C = \frac{IN}{60} = \frac{2\pi N R_{max}}{\vartheta_h\,\tau \times 30\,c}. \quad \ldots\ldots \quad (9)$$

Inserting the value of R_{max} given by (4), with $H = \sqrt{\lambda/\vartheta_h\vartheta_v}$, we find:

$$C \propto \frac{N\lambda^{\frac{1}{2}}\,P_0^{\frac{1}{4}}}{\tau\vartheta_h^{\frac{3}{2}}\,\vartheta_v^{\frac{1}{2}}\,P_{min}^{\frac{1}{4}}}. \quad \ldots\ldots \quad (10)$$

We now eliminate from this formula the two variables P_{min} and τ. The minimum detectable signal P_{min} can be put proportional to the bandwidth B, and since $B = (1 \text{ to } 2)/\tau$, we have:

$$P_{min} \propto \frac{1}{\tau}.$$

As is clear from (10), to get the best C it is desirable to make the pulse length τ as short as possible. Since the minimum pulse length corresponds to a certain number of high-frequency oscillations of the magnetron (for both build-up and full power of the pulse) we can obviously say that the minimum obtainable pulse length is approximately proportional to the wavelength:

$$\tau_{min} \propto \lambda \propto \frac{1}{f}.$$

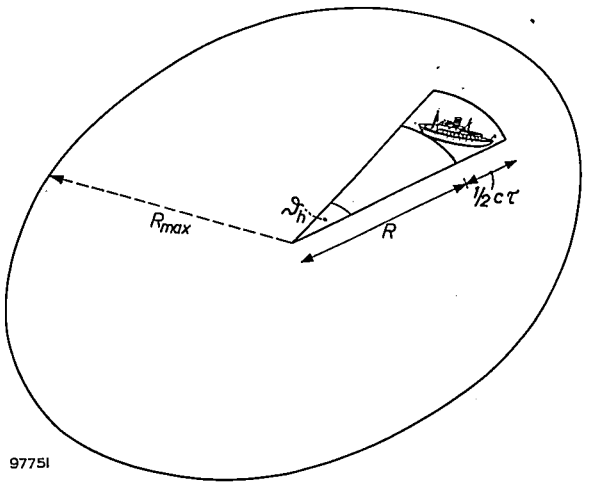


Fig. 5. Determination of the number of objects observabl with a radar system.

This makes the information capacity:

$$C \propto \frac{N f^{\frac{1}{4}} P_0^{\frac{1}{4}}}{\vartheta_h^{\frac{3}{2}} \vartheta_v^{\frac{1}{2}}}. \quad \ldots\ldots\ldots \quad (11)$$

From this we conclude that if we increase the frequency and reduce the antenna dimensions so as to keep to beam-width constant, the information capacity increases in accordance with the fourth root of the frequency.

Assuming constant antenna dimensions, l_h and l_v, we can write the formula somewhat differently. We substitute:

$$\vartheta_h \propto \frac{\lambda}{l_h} \propto \frac{1}{f l_h} \quad \text{and} \quad \vartheta_v \propto \frac{\lambda}{l_v} \propto \frac{1}{f l_v}.$$

The information capacity formula then becomes:

$$C \propto N l_h^{\frac{3}{2}} l_v^{\frac{1}{2}} f^{\frac{9}{4}} P_0^{\frac{1}{4}}. \quad \ldots\ldots \quad (12)$$

This leads to a second conclusion, namely that for constant antenna dimensions the information capacity increases with the frequency to the power of nine fourths.

We may be inclined to conclude that increasing the speed of revolution N of the antenna would increase the information capacity. As indicated by formulae (5), (6) and (7), an increase of N is associated with an increase in the minimum required pulse repetition frequency F_{min}, whereas it is necessary to remain under the maximum value F_{max} corresponding to the maximum average power output of the magnetron. A higher C as a result of increasing N is thus only possible within narrow limits. With regard to (12) it must also be remembered that a higher frequency implies according to (1) a narrower antenna beam, which in turn implies in accordance with (7) that the pulse repetition frequency must also be increased.

Choice of wavelength in relation to atmospheric conditions

Not every wavelength in the millimetre range is suitable for the purposes of radar. The reason is that, owing to atmospheric absorption, attenuation in this range is much greater than in the centimetre range. Closer examination shows that there are various reasons for this, such as absorption by oxygen and water vapour, and attenuation by particles of water in the form of rain, mist, clouds, etc.

As regards the first of these causes, oxygen is found to have two absorption bands, viz. at wavelengths of 5 mm and 2.5 mm; water vapour shows absorption bands at various wavelengths in the mm range and at much shorter wavelengths. The absorption bands of interest to us are at $\lambda = 13.4$ mm and $\lambda = 1.64$ mm. This follows from investigations by J. H. van Vleck [4]), who studied the mechanism of this attenuation and calculated the numerical values as a function of wavelength. The curves in *fig. 6* give as a function of wavelength the attenuation suffered by radio waves in the atmosphere as a result of oxygen and water vapour. The values were calculated for a pressure of 76 cm Hg and a temperature of 20 °C. The relative humidity of the water vapour is taken as 100%. (The attenuation is of course less when the degree of humidity is lower.)

When the atmosphere contains water drops in the form of rain, mist and the like, there are two causes of attenuation:

a) energy is lost by *scattering* of the beam, and

b) energy is *absorbed* and converted into heat (dielectric losses).

In the case of water drops of very small diameter (fog, clouds) the attenuation is due mainly to absorption rather than scattering. The attenuation is then proportional to the quantity of water per m³. The attenuation due to these causes has been calculated by J. W. Ryde and D. Ryde [4]). *Fig. 7* shows the attenuation curves for different degrees of fog and rain as a function of wavelength.

Curve *3* in fig. 6, which gives the total attenuation without water drops, is seen to have a minimum at a wavelength of 8.2 mm. Since our radar equipment

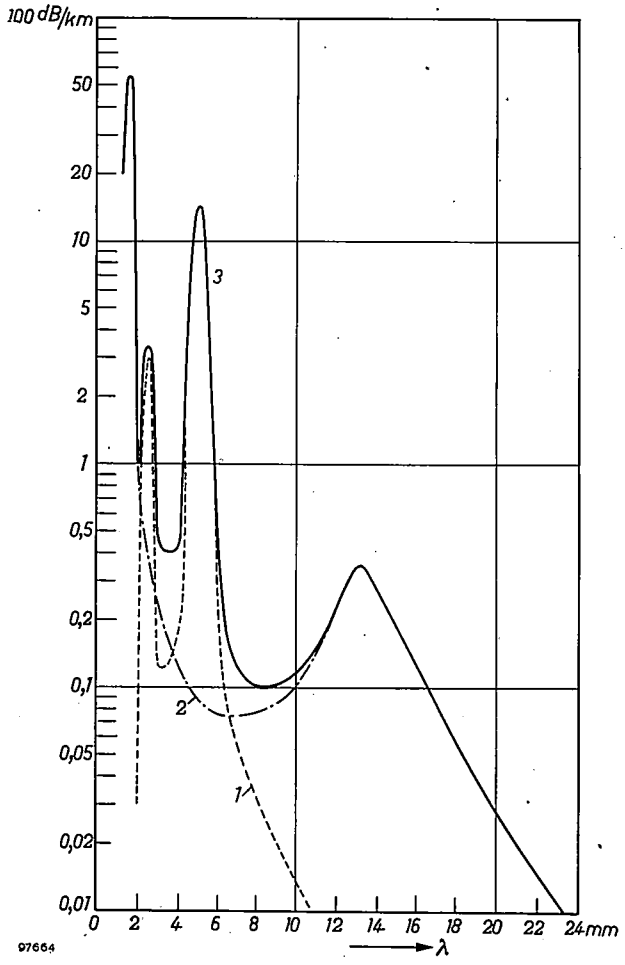


Fig. 6. Attenuation of radio waves by oxygen and water vapour in the atmosphere, as a function of wavelength. *1* = oxygen, *2* = water vapour, *3* = total attenuation (sum of *1* and *2*).

[4]) See e.g. D. E. Kerr, Propagation of short radio waves, Radiation Laboratory Series, No. 13, McGraw-Hill, New York 1951, Chapter 8.

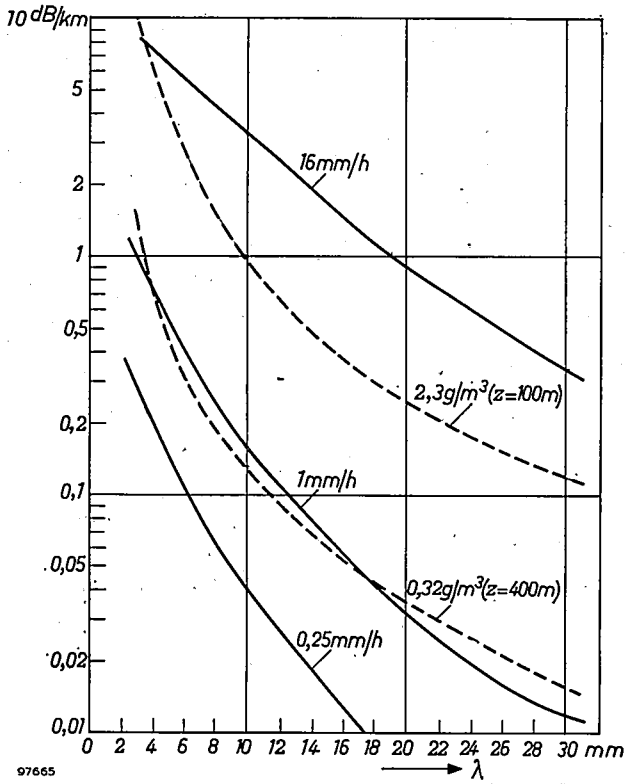


Fig. 7. Attenuation of radio waves caused by different degrees of rain and fog as a function of wavelength. The fully-drawn curves relate respectively to heavy rain, light rain and drizzle; the dashed curves relate to fog, grammes of water per m^3 and the visibility z in metres being indicated.

is obviously to be used in rain and fog, the attenuation caused by water drops must also be taken into account, so that the minimum shifts to $\lambda = 8.6$ mm.

The fact that the minimum attenuation per kilometre is greater than in the centimetre range (e.g. 3 cm) is not a serious objection, the 8 mm radar installation being intended for short distances where the total attenuation is relatively small.

Description of the 8 GR 250 8 mm radar installation

The layout of the installation is illustrated in *fig. 8*. The transmitter, the receiver and the modulator plus drive stage are contained in a water-tight casing (*TMR*). On this casing are mounted the antennae and their driving mechanism. The antennae are rotated at a speed of 40 revs per minute. As already indicated in fig. 1, separate antennae are used in the 8 GR 250 for transmitting and receiving. In general radar practice, and even with 3 cm radar [5]) this is not usual, one antenna being used to which the transmitter and the receiver are alternately switched by a so-called duplexer (see *fig. 9*).

[5]) See Philips tech. Rev. **20**, 349-353, 1958/59 (No. 12).

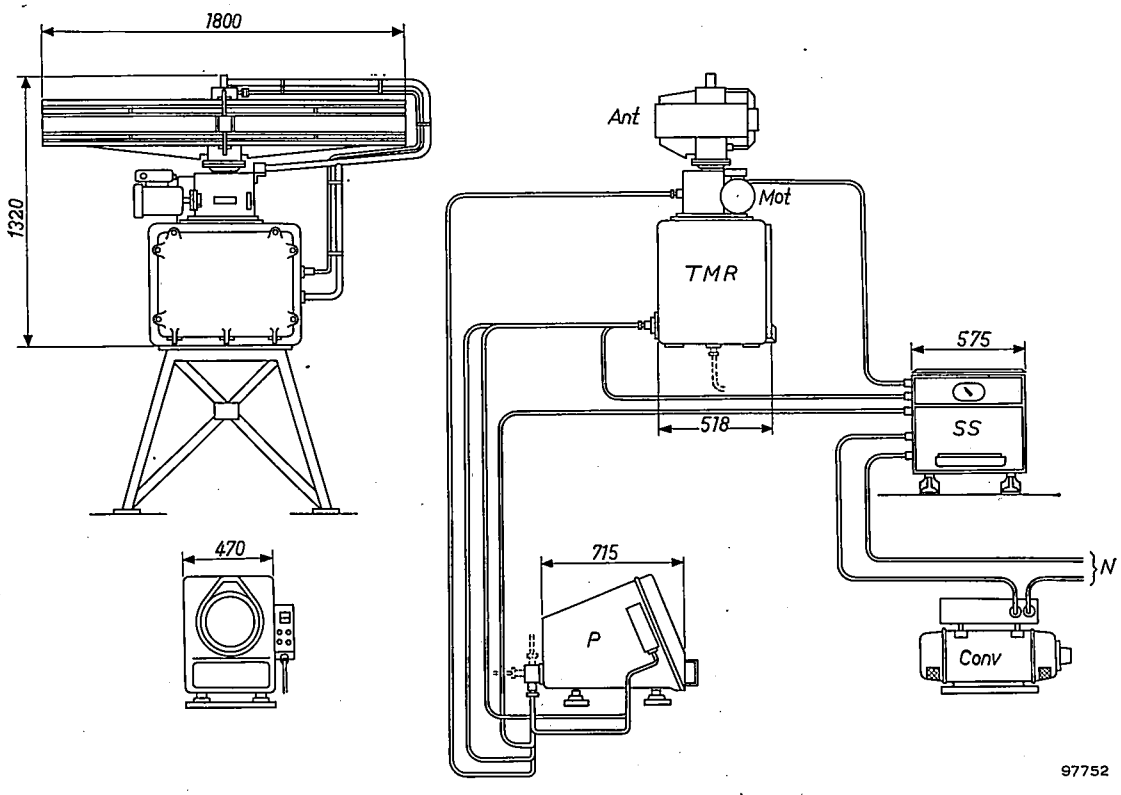


Fig. 8. Schematic layout of Philips 8 GR 250 8 mm radar installation. *TMR* water-tight case containing transmitter, modulator and receiver. Above the case are the two antennae *Ant* with drive motor *Mot*. *P* is the PPI unit. *SS* contains the supply circuits and the synchronizing pulse generator. *Conv* is a rotary converter supplying 120 V, 400 c/s. *N* mains connections. Front views of the *TMR* and *P* sections are shown on the left. Dimensions in millimetres.

A duplexer consists of two gas-filled tubes mounted in a special waveguide circuit. These open the transmitter channel and close the receiver channel while the pulse is being sent out, and reverse the process

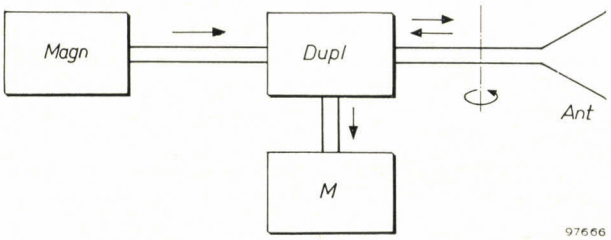


Fig. 9. Waveguide circuit with duplexer *Dupl*, as in a 3 cm radar installation using a single antenna for both transmitting and receiving. *Magn* = magnetron, *M* = mixer in receiver.

in the period between two transmitted pulses. The extinction time of these tubes is too long, however, to allow objects at very short distances to be observed, the reception channels still being closed upon the arrival of such early reflections. This is of course most undesirable for a short-range radar.

The antennae are two parabolic reflectors of the "cheese" type. Crosstalk between the receiving and transmitting antennae can be minimized by mounting the antennae at a specific distance apart. The result is that with relatively little trouble the minimum distance at which an object is seen can be made very small: one can even see on the PPI the feeder waveguide for the upper antenna (which inevitably comes into the beam because of the free rotation of the antennae).

The indicator unit (*P*) is shown bottom left in fig. 8. It consists of the cathode-ray tube with PPI screen, power pack, and the control panel on which, among other things, the radius of the swept area seen on the screen can be adjusted in steps. These radii are given in the table (next page); they differ for the two models mentioned at the beginning of this article, intended respectively for shipboard use or for airfield use.

On the right in fig. 8 can be seen the main power pack (*SS*), which also contains the synchronizing-pulse generator that determines the pulse repetition frequency of 5000 c/s. Below this is the rotary converter, which provides the supply voltage (120 V, 400 c/s).

The table below gives the principal data for the whole installation.

Turning again to the repetition frequency of 5000 c/s, the following may be noted. If we insert in formula (7) the appropriate values from the table, we find the minimum pulse repetition frequency F_{min} to be 800 c/s. This means that every object is scanned by about six pulses at a time, which is a reasonable number for obtaining a favourable signal-to-noise ratio. The maximum average power output of the magnetron is 7.5 W; from formula (6)

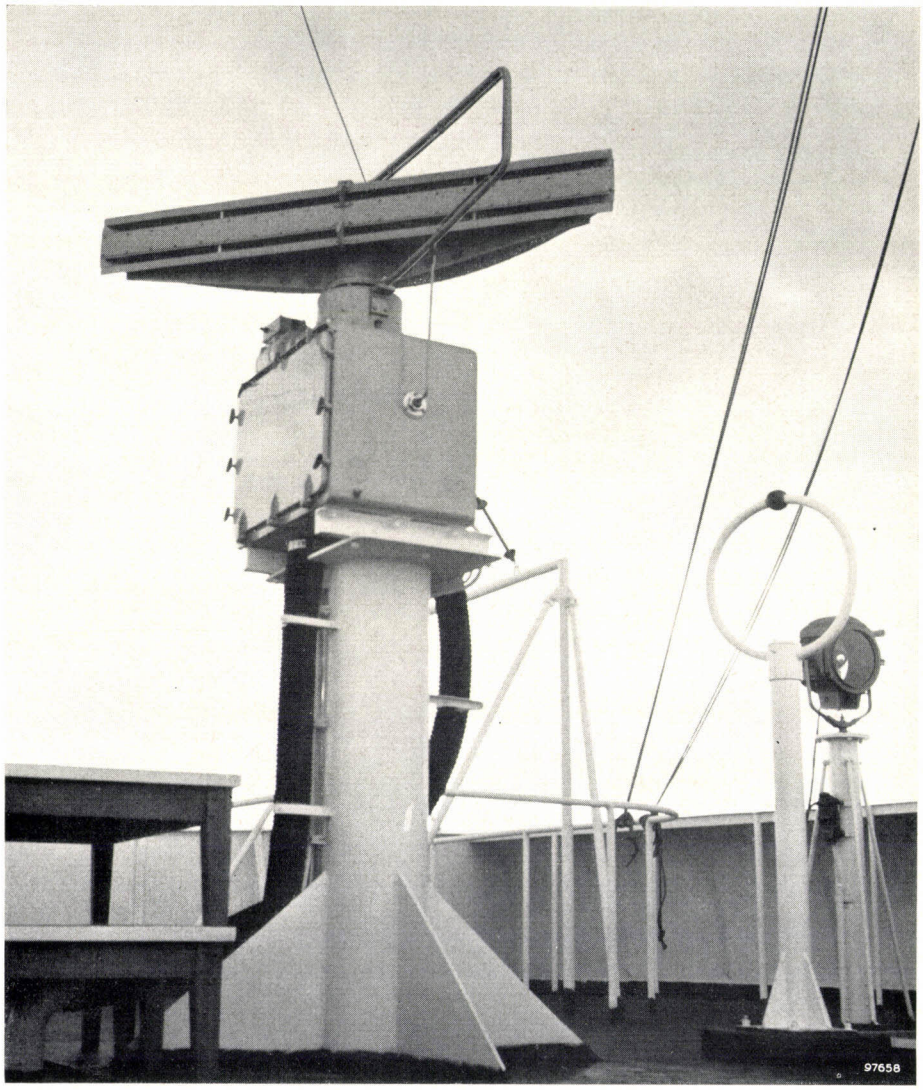


Fig. 10. *TRM* unit and antennae. The loop round the antennae is the feeder waveguide for the upper antenna.

Frequency band	34 512 - 35 208 Mc/s
Pulse length	0.02 μsec
Peak pulse power	25 kW
Pulse repetition frequency	5000 c/s
Intermediate frequency	90 Mc/s
IF bandwidth	50 Mc/s
Beam width of antenna { horizontal	0.3°
Beam width of antenna { vertical	17°
Speed of antenna rotation	40 r.p.m.
PPI	30 cm diameter
Radii of areas displayed on screen: A.S.M.I. type 8 GR 250/00	0.5 - 1 - 1.5 - 3 - 5 - 10 km (550 - 1100 - 1650 - 3300 - 5500 - 11000 yds approx.)
"Ship-shape" type 8 GR 250/01	0.3 - 0.5 - 1 - 2 - 3 - 5 nautical miles
Supply voltage and frequency	120 V, 400 c/s

it therefore follows that the maximum pulse repetition frequency F_{max} is 15 000 c/s. The pulse repetition frequency of 5000 c/s is thus sufficiently far below this value.

Fig. 10 shows a photograph of the watertight casing in which the transmitter, receiver and modulator are mounted. The whole unit is installed on the upper deck of a ship. The slots of the transmitting and receiving antennae above the casing are clearly visible in the photograph. The receiving antenna is the upper one.

approximately 15 kV; the magnetron draws a peak current of about 12 A. The modulator is designed to give the most favourable possible efficiency. The operation of the circuits involved will be discussed with reference to the diagram in *fig. 11*.

Our requirements for the modulator are the following.

1) We need rectangular pulses of the right shape and duration;

2) These must be of high voltage and considerable power;

3) The power must be stored in a reservoir which can be charged up again in the interval between two pulses.

These requirements are fulfilled here by two circuits: the driver stage, which generates the pulses, and the modulator proper, which converts these into pulses of high voltage and power.

As is well known, a good rectangular pulse can be achieved by taking a transmission line of specific length, charging it to a certain voltage and discharging it through a resistor. If the line is lossless and terminated by its characteristic impedance, this discharge will cause a constant current to flow through the resistor for a time that depends on the length of the line. This is the way in which the pulses are generated in the driver stage (see fig. 11). The

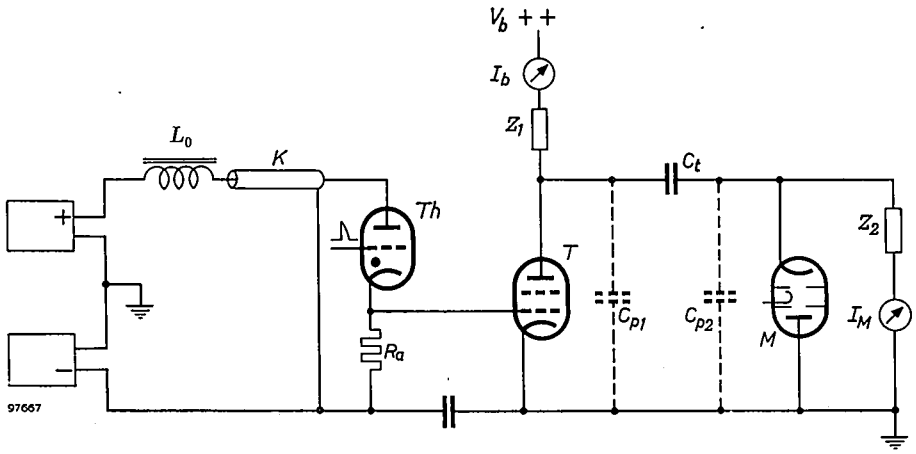


Fig. 11. Simplified schematic diagram of modulator and driver stage. M = magnetron, C_t = storage capacitor, T = tetrode, which functions as a short-circuiting switch, Th = thyratron, which connects the resistance R_a to the pulse-forming network (coaxial line) K at the appropriate moments, L_0 = charging choke, Z_2 = blocking impedance for magnetron filament supply, C_{p1} and C_{p2} = stray capacitances, responsible for much of the dissipation.

Some circuit details

Modulator

The Philips magnetron type 7093 [6]) used in the installation is modulated which voltage pulses of

transmission line K is in this case a simple coaxial cable. This is possible because the pulse length is so short. If it were longer the cable would have to be very long indeed, and in that case use would preferably be made of an artificial line for which the velocity of propagation of electromagnetic waves is smaller.

The thyratron Th is the switch connecting the line K to the resistor R_a over which the voltage

[6]) This magnetron is described by J. Verweel and G. H. Plantinga, A range of pulsed magnetrons for centimetre and millimetre waves, Philips tech. Rev. 21, 1-9, 1959/60.

pulses appear. In the intervals between pulses, K is again charged up gradually via the choke L_0. At the appropriate moment the synchronizing-pulse generator (fig. 1), which determines the pulse repetition frequency, supplies a signal to the grid of thyratron *Th*. This is thereby made conducting and K discharges via R_a. This voltage pulse arrives on the control grid of the pulse tetrode T in the modulator circuit, causing this tube to pass a high current and the anode voltage to drop by about 15 kV until it bottoms to a value E_r. This negative voltage pulse is applied via capacitor C_t to the cathode of the magnetron M, which is thus triggered into oscillation. Since T must not conduct between the magnetron pulses, the supply circuit for the driver stage is so designed as to keep the control grid of T under constant negative bias. This is to be seen from fig. 11 (remembering that the thyratron is kept non-conducting at that moment). In connection with various protective devices is was found necessary to divide the power pack into two parts, indicated in the figure.

A particular feature of the modulator stage, which is basically of conventional design, is that the stray capacitances are kept as small as possible. The losses in the modulator occur in the tetrode T, and as a result of the repeated charging and discharging of the stray capacitances C_{p1} and C_{p2}. When a capacitor is charged from V_1 to V_2 an energy of $\frac{1}{2}C(V_2-V_1)^2$ is lost in the charging resistance, whilst an energy $\frac{1}{2}C(V_2^2-V_1^2)$ is stored in the capacitor, and this is also lost upon discharge. Since the pulse length is so short (0.02 μsec) the latter losses are important. We shall deal with this presently, but we shall first complete the description of the circuit in fig. 11.

The impedance Z_1 consists of a coil and a resistor in series; the object of this impedance is to protect the high-tension supply circuit (++) against short-circuiting during the pulse, and it also functions as the charging resistor for C_t. This capacitor is the reservoir earlier mentioned, which supplies the current to the magnetron during the pulse. This should have a minimum value such that at the end of the pulse the voltage over C_t has only dropped by a small fraction, is spite of the high current drawn by the magnetron. If, however, C_t is made larger than this minimum value, its capacitance with respect to earth also becomes greater. This earth capacitance partly contributes to the stray capacitance C_{p1} which also comprises the output capacitance of T and the wiring capacitance, and partly to the stray capacitance C_{p2}, which includes the input capacitance of the magnetron and the wiring and earth capacitance of Z_2. The impedanze Z_2 contains a coil and a charging resistor, which returns the magnetron side of C_t to earth potential after the pulse; the coil constitutes a blocking impedance in the filament supply lines to the magnetron, its purpose being to prevent the voltage pulse leaking away via these lines and via the earth capacitance of the filament current transformer. This circuit offers virtually no impedance to the filament current.

We shall now analyse the effect of the stray capacitances C_{p1} and C_{p2} more exactly and calculate the efficiency. We start from the quiescent state of the modulator. The high tension V_b is on the anode of T and on one side of C_t and C_{p1}. Both sides of the magnetron M and of C_{p2} and Z_2 are at earth potential. The tetrode T is cut off. As soon as tetrode T becomes conducting, C_{p1} discharges through T until the voltage has dropped to E_r, which is the bottomed voltage across T. In the same time interval the voltage across M increases to $-(V_b - E_r)$, which also applies to C_{p2}. The charge which C_{p2} thereby acquires is given up by C_t. During the magnetron pulse both C_t and C_{p2} lose part of their charge to the magnetron, but only a very small part, as already remarked. The voltage across M and C_{p2} therefore remains approximately equal to $-(V_b - E_r)$ during the pulse. After the magnetron pulse, C_t is charged and C_{p2} is discharged. It follows from the foregoing that the potential of C_{p1} is alternately V_b (between the pulses) and E_r (during each pulse). The corresponding values for C_{p2} are 0 and $-(V_b - E_r)$, if we disregard the insignificant role played by C_{p2} as storage capacitor during the magnetron pulse. We said earlier that this charging and discharging of capacitors costs a certain amount of energy. C_{p1} takes from the high-tension supply from pulse to pulse $C_{p1}V_b(V_b - E_r)$, which is a pure loss. In the same period C_{p2} receives $C_{p2}(V_b - E_r)^2$. Approximately, therefore, we can regard C_{p1} and C_{p2} as a single capacitance C_p. The efficiency of the modulator is now given by the ratio of the input energy to the magnetron for one pulse to the total energy dissipated during the pulse and between two pulses:

$$\eta \approx \frac{(V_b - E_r)I_{tm}\tau}{C_pV_b(V_b-E_r)+E_rI_{tm}\tau+(V_b-E_r)I_{tm}\tau}, \quad (13)$$

where I_{tm} is the peak current of the magnetron, τ is the pulse length and $E_rI_{tm}\tau$ is the loss in the modulator tube T. If we insert the following values: $V_b = 16.5$ kV, $E_r = 1.5$ kV, $\tau = 0.02$ μsec, $C_p = 23$ pF, and $I_{tm} = 12$ A, we find $\eta = 40\%$. The energy dissipated per pulse is proportioned as follows:

dissipation in the magnetron 3.6 mJ,
dissipation in the tetrode 0.36 mJ,
loss due to stray capacitances 5.7 mJ.

This clearly indicates the relatively important role played by the stray capacitances, and which is entirely a consequence of the short pulse length τ.

The efficiency can be checked during operation by noting the meter reading for the average magnetron current I_M and the average current I_b of the high tension supply.

Pulse correcting circuit

The echo pulse picked up by the antenna may suffer distortion as a result of reflection, the effect being that the pulse is elongated, i.e. it acquires a "tail". (The smaller the length of the wave train of the magnetron pulse with respect to the reflecting object, the greater the likelihood of tail formation.) This is obviously undesirable since it reduces the resolution of the radar system. For this reason the 8 GR 250 contains a pulse-correcting circuit, which can be put into operation when needed. The correction is obtained by adding to the elongated pulse a pulse of the same shape, but of opposite phase and of amplitude equal to the mean amplitude of the tail, delayed to an extent corresponding to the pulse length τ; see *fig. 12*. The result is a pulse whose tail is largely eliminated.

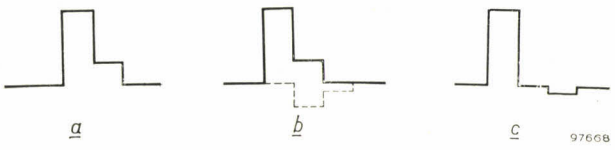


Fig. 12. *a*) Echo pulse with (idealized) tail. *b*) Delayed and inverted pulse added to (*a*). *c*) Corrected pulse.

This correction is effected by means of a four-terminal network as shown in *fig. 13*. (If necessary the procedure can be repeated with similar networks.) The network is placed in the coaxial video-line, of characteristic impedance ζ, and is composed of a series and a parallel resistance of magnitude ζ. In parallel with each of these two resistors is another coaxial line of characteristic impedance ζ and electrical length $l = c\tau/2$, and terminated respectively by resistors $R' = \zeta^2/R$ and R. When a voltage $U_1(t)$ is applied to the input terminals, we can write for the output voltage $U_2(t)$:

$$U_2(t) = \tfrac{1}{2}\{U_1(t) + kU_1(t-\tau)\}, \quad . \quad . \quad (14)$$

where $k = (R - \zeta)/(R + \zeta)$. A suitable choice of k, that is of R, produces the phenomenon illustrated in fig. 12, for the formula indicates that we add to the original pulse a delayed pulse of the required polarity.

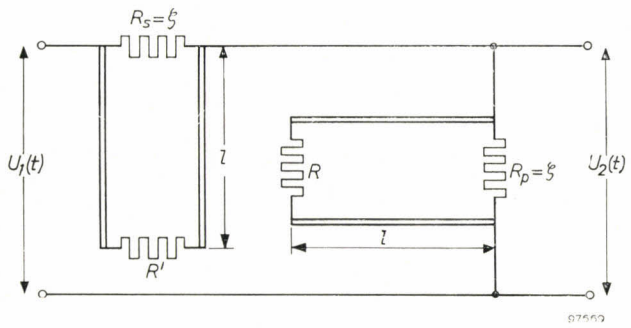


Fig. 13. Pulse-correcting network. Parallel with the series resistor R_s is a coaxial line of length l terminated by R'; parallel with R_p is likewise a coaxial line of length l, terminated by R.

Fig. 14. Radar display of a London dock area, taken with the 8 GR 250. The diameter of the display is adjusted for a coverage of half a nautical mile.

The formula given here can be simply derived with the aid of the theory of long transmission lines. In doing so we disregard the line losses. The input impedance is ζ. If the voltage on the input terminals be $U = U_1 e^{j\omega t}$, the voltage at the output terminals is:

$$U_2 = U_1 e^{j\omega t} \frac{Z}{Z+\zeta},$$

where

$$Z = \zeta \frac{R + j\zeta \tan al}{\zeta + jR \tan al}$$

and $a = \omega/c$, $al = \omega l/c = \omega\tau/2$.

The form $Z/(Z + \zeta)$ can be written:

$$\frac{Z}{Z+\zeta} = \frac{1}{2}\left(1 + \frac{R-\zeta}{R+\zeta} e^{-j2al}\right).$$

Putting $(R - \zeta)/(R + \zeta) = k$, which can be either positive or negative, we find:

$$U_2 = \frac{1}{2}\left(U_1 e^{j\omega t} + kU_1 e^{j\omega(t-\tau)}\right).$$

Fig. 15. Radar display of apron at Schiphol airport, taken with the 8 GR 250. Screen set to $\frac{1}{2}$ km diameter ($\sim$ 550 yard).

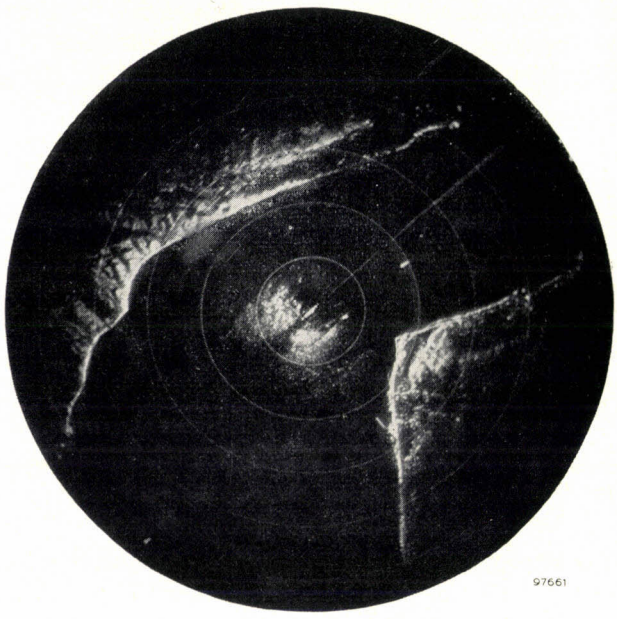


Fig. 16. Radar display of Straits of Messina, taken with the 8 GR 250; screen coverage set to approximately 5 km diameter ($\sim$ 2.7 nautical miles).

Thus, if we apply a sinusoidal voltage to the input terminals we obtain at the output terminals half the sum of the input voltage U, increased by a fraction k of U which is delayed by a time τ (the pulse length) and is independent of the angular frequency ω. For an arbitrary voltage $U_1(t)$ each Fourier component of its spectrum will therefore suffer the same delay, and we obtain formula (14).

Performance of the 8 mm radar installation

The 8 GR 250 installation has been tried out in various circumstances and at many different locations. Some of the results are illustrated in photographs of the radar screen shown in this article.

Figs 2 and 3 show respectively the map and the radar display of Ymuiden Harbour. The opened lock-gates and other details of the docks are distinguishable on the radar screen [7]).

Fig. 14 shows the radar display of a part of the London docks. The area covered by the screen has a radius of about 900 metres. Here again, the high resolution of the apparatus is clearly brought out; one can see on the right, for example, the quays with ships alongside. At the centre even the shape of the s.s. Amstelstroom, on which the radar equipment was installed, can be roughly distinguished.

Fig. 15 shows a radar display of the apron at Schiphol airport (coverage radius 500 m). The shapes of some aircraft are clearly visible, and a ship can also be seen in the Ringvaart canal.

Finally, *fig. 16* was obtained on board a ship passing through the Straits of Messina. It can be seen that, in spite of the attenuation of 8 mm waves in the atmosphere, a good picture is also obtained of a larger area (radius of coverage about 5 km).

Summary. When radar is used for very short ranges, for example in harbours, it is most desirable to be able to distinguish the shape as well as the distance of the surrounding objects. This means that the radar equipment must posses a high resolution. This can be achieved by using the shortest practicable wavelength, which makes it possible to obtain a pulse of extremely short duration and a beam of very small angular width. The radar equipment described, here Philips 8 GR 250, operates on a wavelength of 8.6 mm, which comes within a frequency region at which the atmospheric attenuation of millimetre radio waves has a minimum. Some general theoretical considerations are given with regard to the effect of wavelength on such properties as resolution, range and information capacity. An appropriate choice of such quantities as antenna dimensions (horizontal beam-width 0.3°), pulse length (0.02 μsec) etc., leads to a design in which the best compromise is achieved between the radial and tangential resolution of the radar system and the resolution of the PPI. The narrow beam entails the use of a high pulse repetition frequency. In the design of the modulator it was necessary to keep the stray capacitances as small as possible, since, owing to the short pulse length, these give rise to losses that are considerable compared with the power taken up by the magnetron. To ensure that the received pulse has a good shape under all conditions, a pulse-correcting network is incorporated in the video section of the receiver. Some photos of PPI displays illustrate the impressive results obtained with this equipment.

[7]) It is instructive to compare this with a 3 cm radar picture of Ymuiden as reproduced in Philips tech. Rev. **14**, 95, 1952/53.

A METAL-CERAMIC DISC-SEAL TRIODE FOR FREQUENCIES UP TO 6000 Mc/s

by E. MENTZEL*) and H. STIETZEL*)

621.385.3.029.64

Since the completion of the EC 59 disc-seal triode [1]), work has proceeded on the development of a partly ceramic version of this microwave valve. The perfecting of the technique of bonding metals to ceramics has enabled very low loss resonant cavities to be produced. Moreover the use of a ceramic instead of glass makes it possible to increase the power output of grid-controlled valves, and also to raise the upper frequency limit — in so far as the latter is governed by valve geometry. On these lines a new disc-seal triode has now been made, provisionally designated as type OZ 92 and suitable for operation at frequencies up to 6000 Mc/s (*fig. 1*). The anode voltage is 500 V and the anode dissipation 125 W. The valve can be cooled either with water or with air. The power output at 5000 Mc/s is more than 10 W. The valve is not yet commercially available.

The requirements imposed on grid-controlled microwave valves are broadly: small inter-electrode spacings (to minimize transit-time effects); high transconductance; low valve capacitances (implying high specific loading of electrodes); and low coupling between grid-anode and grid-cathode circuits by optimum design of the grid (to achieve a high gain).

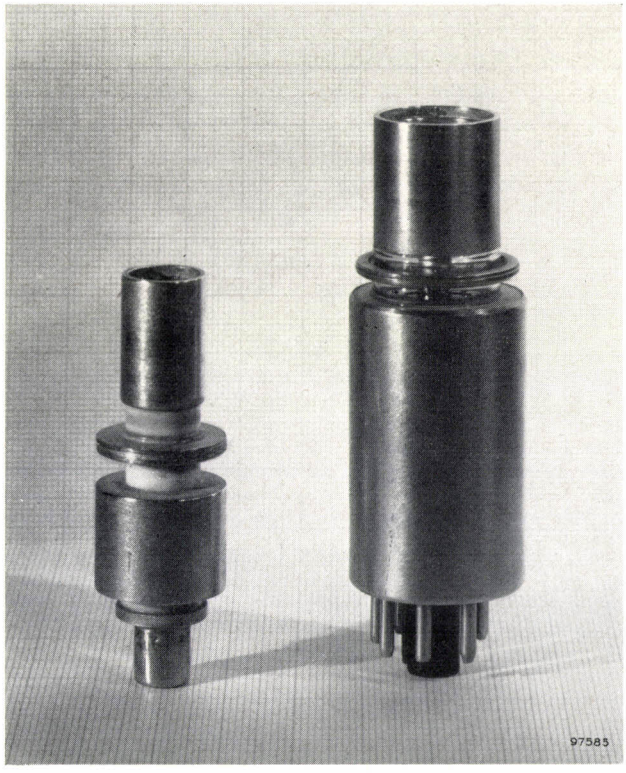


Fig. 1. Experimental disc-seal triode type OZ 92 with metal-ceramic wall (left) and disc-seal triode type EC 59 with metal-glass wall (right).

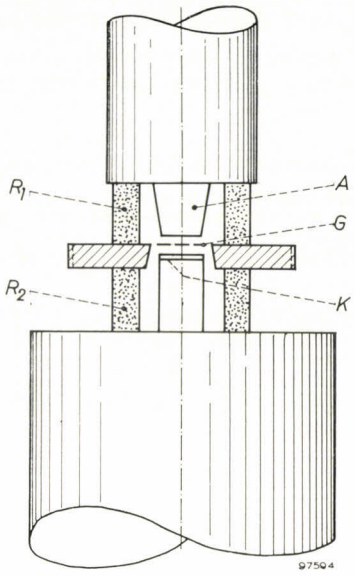


Fig. 2. Schematic cross-section of disc-seal triode OZ 92. K cathode, G grid, A anode, R_1 and R_2 ceramic rings.

Since disc-seal triodes are always mounted in resonant cavities, the shape of the valve wall influences the possibilities of tuning these cavities. The shape given to the valve envelope can therefore modify quite appreciably the upper limit of the useful frequency range.

As a material for valve walls sintered *aluminium oxide* ceramic offers considerable advantages, chiefly because the Al_2O_3-metal seals are capable of high thermal loading and Al_2O_3 has low dielectric losses even at high temperatures. The appropriate metal for the seal is an alloy of iron, nickel and cobalt. The solder used is silver or an alloy of silver and copper. The technology of metalceramic seals is now so far advanced that the kind of joint made — whether a flat or a collar joint — does not essentially affect the vacuum seal or its strength. The sealing technique therefore imposes little restriction on the design of the valve wall.

Fig. 2 shows the configuration adopted for the wall of the OZ 92. The design is based on the following considerations. In the first place it allows the valve to be mounted in conventional sections of waveguide, and in the second place, owing to the low expansion of the anode portion of the wall the electrode clearances retain the precision required even for very high frequencies. Some examples of

*) Development laboratory of the Valvo G.m.b.H. Radio Valve Factory, Hamburg.

1) V. V. Schwab and J. G. van Wijngaarden, The EC 59, a transmitting triode with 10 W output at 4000 Mc/s, Philips tech. Rev. **20**, 225-233, 1958/59 (No. 8).

this configuration are the disc-seal triodes EC 156 and EC 157 [2]), and the EC 59 [1]), already mentioned.

Features of the EC 59 taken over in the OZ 92 are the electrode dimensions and clearances; also the new valve is run under the same D.C. conditions. The frequency limit is thus fixed, in so far as it is determined by electrode clearances and diameters. The frequency limit being fixed, the outside dimensions of the valve follow mainly from two requirements:

1) The resonant cavity should oscillate in the fundamental mode (E_{010} resonance of the capacitatively loaded cylindrical cavity).
2) The surface area of the valve wall where high-frequency current flows must be made as small as possible in relation to the total surface area of the resonant cavity.

The second requirement is a consequence of the fact that, in the walls of a capacitatively loaded E_{010} resonant cavity, only weak currents flow near the axis, so that even if the conductivity of the surface there is poor, high losses are not incurred in this part of the wall; contact resistance losses are also negligible near the axis. Ohmic losses in parts of the cavity walls farther from the axis can be minimized by improving the surface (silver-plating, polishing). The Q of a resonant cavity depends on its form and on the electrical losses. The E_{010} mode being the fundamental mode, the Q of such a resonant system is optimum when the height of the cavity is approximately half the diameter. On the basis of these data we can calculate the geometrical dimensions. At the given values of grid-anode separation (h, see *fig. 3*) and anode diameter (d) we find for 5000 Mc/s the diameter (D) and the height (H) of the cavity as given in the caption to fig. 3.

The ceramic rings for the valve wall have an outside diameter of 11.6 mm, a thickness of 1.6 mm and a height of 6 mm. These dimensions are the result of a compromise between the need to make the valve as small as practicable and the fact that the ceramic material, owing to its high dielectric constant ($\varepsilon_r = 10$), would constitute a considerable capacitative load on the valve if the dimensions were unduly small, thereby lowering the Q of the anode resonant cavity. For the dimensions given the limit frequency of the finished valve wall at the grid-anode side is 6500 Mc/s.

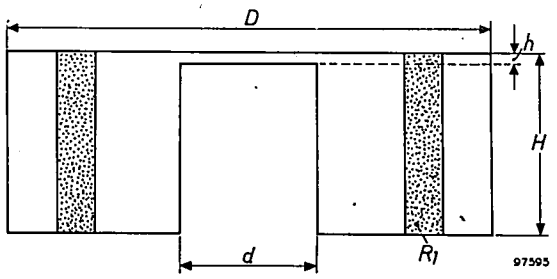


Fig. 3. Schematic cross-section of anode resonant cavity of disc-seal triode with ceramic ring R_1. The Q of the resonant cavity is optimum when $H \approx \frac{1}{2}D$ The following are the dimensions of the 5000 Mc/s cavity of the OZ 92 : $H = 6$ mm, $D = 16$ mm, $h = 0.30$ mm, $d = 4.5$ mm.

To minimize the surface area of the resonant cavity, protruding edges, projecting metal parts and similar irregularities are avoided as far as possible. The high ohmic losses of the Fe-Ni-Co alloy used are eliminated by gold-plating the surface (to a thickness of 3 to 5 μ). As a result of these measures the Q of an unloaded grid-anode cavity of the form shown in fig. 3, with valve walls prepared as above was found to be about 500 at 5000 Mc/s. This is roughly twice the figure obtained with glass-walled valves, and for the conventional applications it means an increase in circuit efficiency of approximately 10%.

The OZ 92 triode consists of four sub-assemblies (*fig. 4a, b, c, d*) with altogether 15 components (corresponding glass-walled valves have some 30 components). All assembly operations are straightforward and can be carried out with the necessary precision by semi-skilled personnel equipped with simple tools and jigs. This simplification of construction and assembly is largely attributable to the good resistance to deformation of the ceramic material in the temperature range below 1000 °C.

The anode assembly (fig. 4*a*) consists of the anode, a ceramic ring and the upper part of the grid disc (the lower part of which belongs to the external cathode assembly). Making the grid disc in two parts simplifies the adjustment of the electrode clearances, the cleaning of internal surfaces of the valve and the finishing of the grid seating. The anode, too, is in two parts, for the practical reason that the Fe-Ni-Co alloy used cannot withstand the high thermal load (1.9 kW/cm²) to which the anode proper is subjected during outgassing on the pump. The anode proper is therefore made of molybdenum. For locating the ceramic ring, the anode and the grid disc are recessed (0.8 mm) which strengthens the structure without significantly increasing the ohmic losses.

Fig. 4*b* shows the external cathode assembly. It comprises the lower part of the grid disc, a second ceramic ring, the cathode sleeve, a third ceramic ring and the outer conductor of the concentric heater-supply system. The metal-ceramic joints in this assembly are similar to that in the anode system. The cathode bush carrying the cathode proper

[2]) Valves EC 156 and EC 157 differ from the older types EC 56 and EC 57 in their cathode, which has a longer life. The EC 57 is described in Philips tech. Rev. **18**, 317, 1956/57.

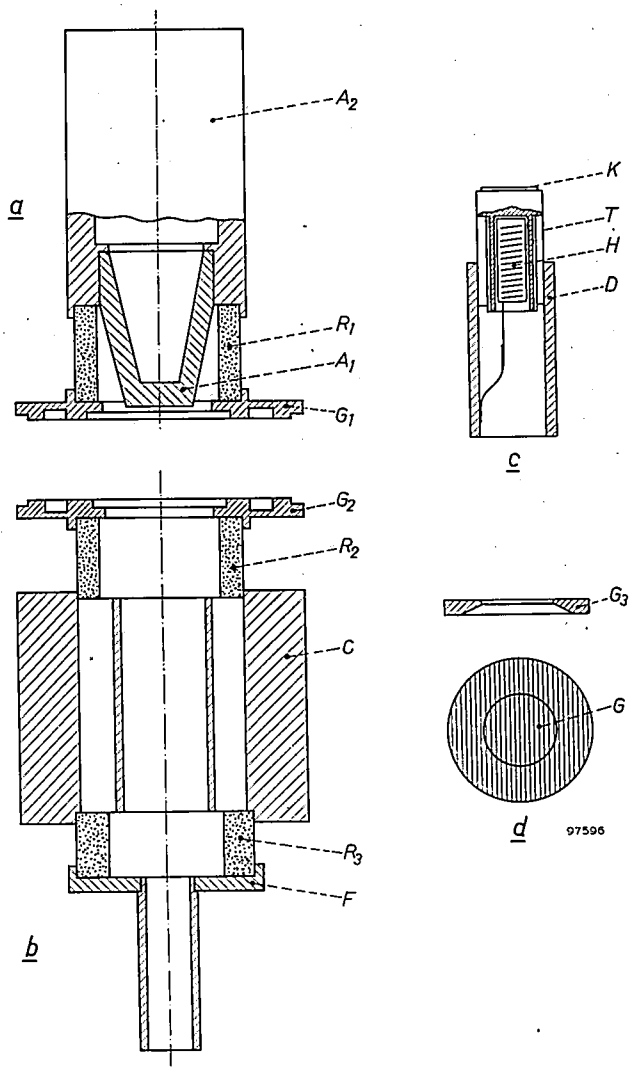


Fig. 4. The four sub-assemblies making up the OZ 92.
a) Anode assembly with anode $A_1 + A_2$, ceramic ring R_1 and upper part G_1 of grid disc. Part A_1 of the anode is of molybdenum, part A_2 is an alloy of iron, nickel and cobalt.
b) External cathode assembly, with lower part of grid disc, ceramic rings R_2 and R_3, cathode sleeve C and heater connection F.
c) Internal cathode assembly, with dispenser cathode K, heater H, tantalum cylinder T and cathode bush D.
d) Grid assembly, with grid G of tungsten wire wound on the ring G_3.

fits into the cathode sleeve (fig. 4*c*). The cathode bush contains the dispenser cathode, the latter being enclosed in a tantalum cylinder which prevents the penetration of evaporated barium into the grid-cathode space and also acts as getter.

The grid assembly can be seen in fig. 4*d*, and consists of the grid-frame wound with fine tungsten wire. The frame is circular and automatically centres the assemblies *a* and *b*. The thickness of the frame determines the spacing between grid and cathode. In the final assembly stage, sub-assemblies *a* and *b* are joined together by high-speed silver-soldering of the two grid disc G_1 and G_2 along their peripheries (speed being necessary to prevent too much slackening in the tension of the tungsten wires).

The D.C. data and other particulars of the OZ 92 are listed in *Table I*. *Fig. 5* shows the I_a-V_g characteristic of an average valve.

Table I. Data of the OZ 92.

Heater voltage V_f . 6.3 V	Transconductance S approx. 17 mA/V
D.C. andode voltage V_a . . . . . 500 V	Amplification factor μ . . . . . approx. 28
D.C. anode current I_a . . . . . 250 mA	Valve capacitances:
D.C. grid bias V_g . . —5 V	C_{gk} = 3.3 pF
Anode dissipation P_a . . . . . 125 W	C_{ag} = 1.9 pF
	C_{ak} = 0.048 pF

Because of the high thermal conductivity of Al_2O_3 and the absence of electrolysis, even at high temperatures, the OZ 92 can be subjected to a very heavy thermal load. The temperature of the metal-ceramic joints, however, must not exceed 250 °C (in the case of the EC 59 the temperature of the glass seal must not rise more than 120 °C above ambient.) This requirement can be met by water-cooling and also by air-cooling. *Fig. 6* shows the temperature of a seal above the ambient as a function of anode

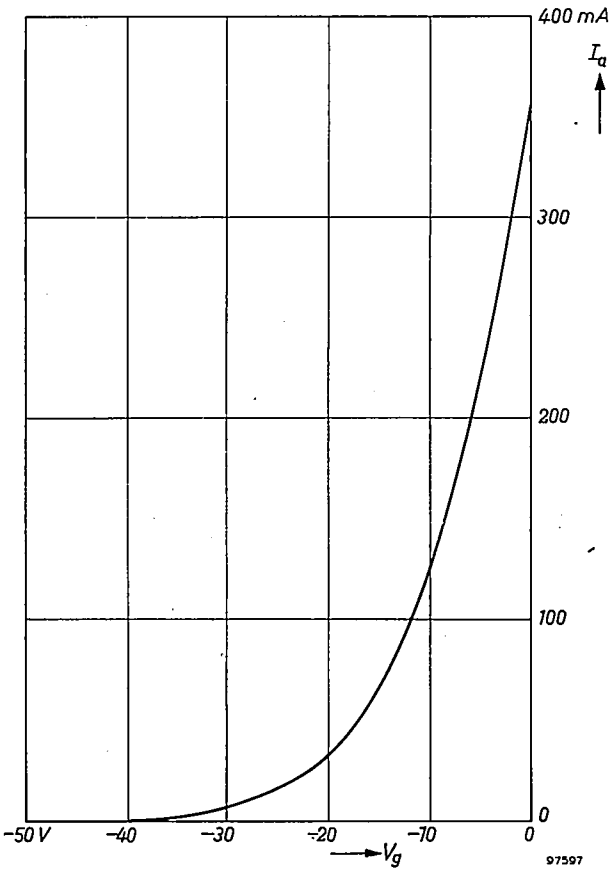


Fig. 5. D.C. characteristic of the OZ 92: anode current I_a as a function of grid voltage V_g, at 500 V anode voltage and 6.3 V heater voltage.

N. V. PHILIPS' GLOEILAMPENFABRIEKEN

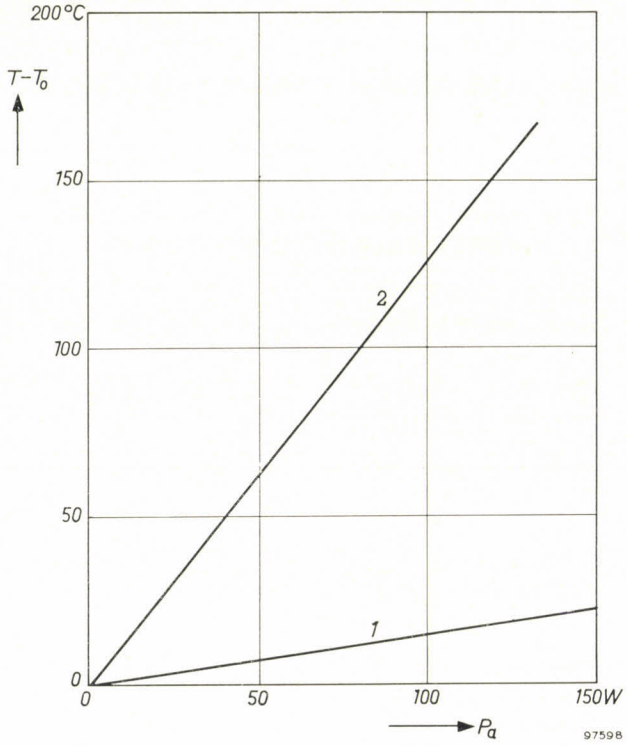


Fig. 6. Temperature difference $T—T_0$ between anode metal-ceramic seal and ambient, as a function of anode dissipation P_a. Curve *1*: water cooling (0.5 l/min); curve *2*: air cooling (40 l/min).

dissipation using experimental cooling systems. The temperature was measured in the immediate proximity of the metal-ceramic seal on the anode.

Another favourable property of the valve is its insensitivity to rapid fluctuations of temperature. This is due to the fact that the expansion coefficients of Al_2O_3 and the Fe-Ni-Co alloy used are virtually identical.

The valve can operate either in amplifier or oscillator circuits up to 6000 Mc/s. *Fig. 7* shows a test set-up with the OZ 92 as an oscillator. The oscillator is of the re-entrant type (*fig. 8*), the frequency being changed by inserting another resonant cavity in the output circuit. Some results obtained with this oscillator are given in *Table II*.

Table II. Experimental results obtained with an OZ 92 in an oscillator circuit, with an anode voltage of 500 V.

Frequency f . . . . . (Mc/s)	4740	4980	4990	5470	5530
Power output P_o . . . (W)	7.1	12.4	13.2	4.8	6.9
Anode current I_a . . (mA)	235	255	320	196	290
Grid bias V_g . . . . . (V)	–25	–17.5	–28	–31	–5
Grid current I_g . . . . (mA)	30	20	33	19	23

The power output of the oscillator falls off at both sides of the frequency band around 5000 Mc/s. Towards higher frequencies the fall-off is presumably due to valve properties, and towards lower frequencies it is attributable to the properties of the oscillator.

The measured outputs in the oscillator circuit — viz. 13 W at 5000 Mc/s and 7 W at 5500 Mc/s, roughly corresponding to efficiencies of 9% and 5%, respectively — prove that the OZ 92 still operates quite satisfactorily at these frequencies.

However, the valve is not primarily intended for use as an oscillator but, like the EC 59, as a *power*

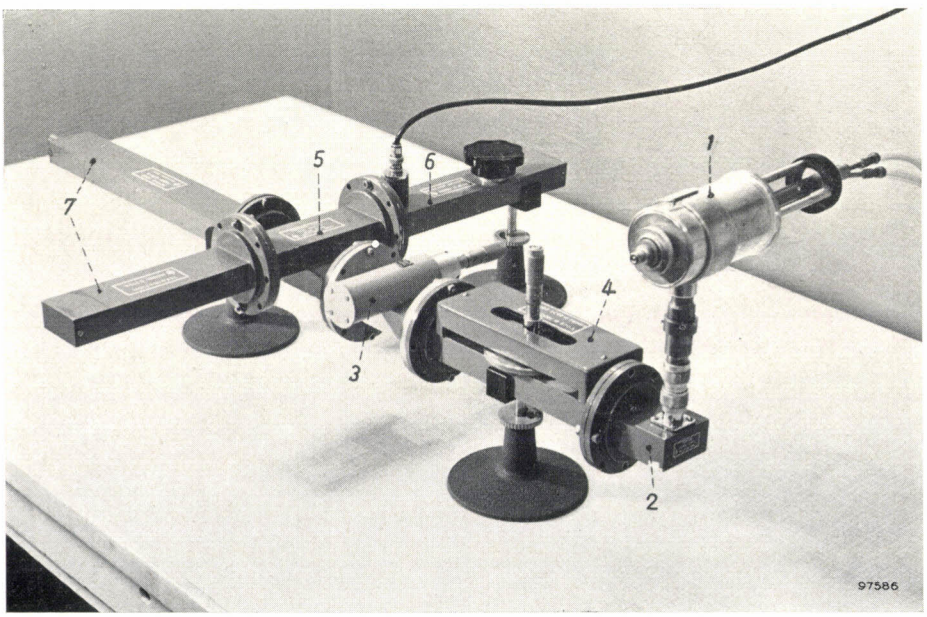


Fig. 7. Test set-up with the OZ 92 operating as an oscillator. The photograph shows the oscillator *1* (cf. fig. 8) and associated waveguide components for measuring purposes: transition section *2*, frequency meter *3*, matching section *4*, cross coupling *5*, thermistor termination *6* and matching impedance *7* for the G band (3950-5850 Mc/s).

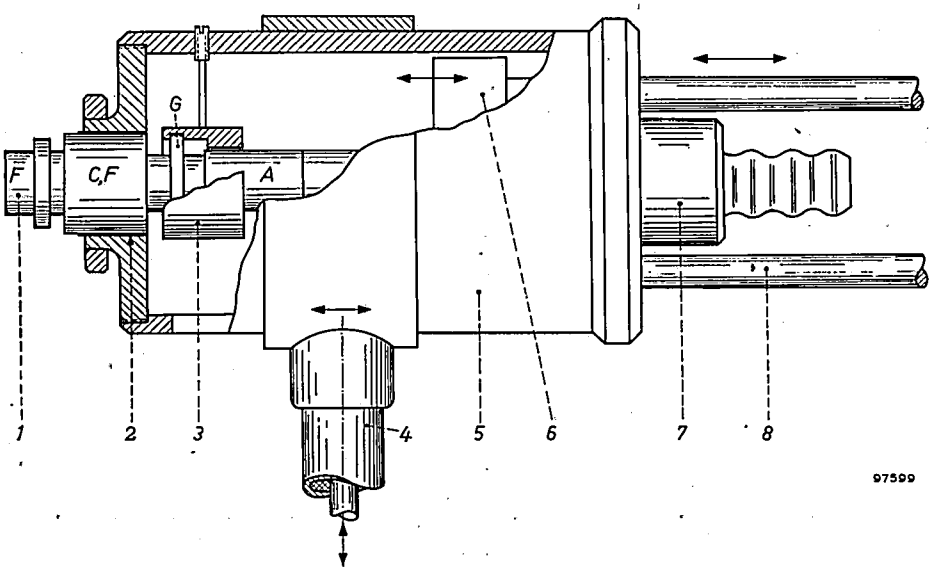


Fig. 8. Oscillator with re-entrant type of cavity, using the OZ 92. *1* disc-seal triode OZ 92. *2* decoupling capacitor. *3* cylindrical bush forming anode resonant cavity. *4* output probe movable axially and radially. *5* outer resonant cavity, also forming the grid connection. This cavity is coupled to the anode resonant cavity by the annular gap between *A* and *3*. *6* tuning plunger for outer cavity (not in contact with *5*). *7* coolant duct, also forming the anode connection. *8* guide rods of tuning plunger.

amplifier. It might, for example, be used with advantage in the G band (3950 - 5950 Mc/s). It behaves electrically rather like the EC 59, but because of its different shape it is not directly interchangeable with it. A simple amplifier designed for inclusion in a waveguide system can be seen in *fig. 9*.

Summarizing it may be said that the advantages of the OZ 92 are due to its walls of aluminium oxide ceramic, the properties of which are superior to those of glass. The advantages of the OZ 92 are:

1) Higher permissible thermal loading, owing to better heat conduction and rigidity of the ceramic. No electrolysis in the ceramic, even at high operating temperatures.

2) The low dielectric losses of the ceramic result in resonant cavities of higher Q.

3) The resistance of the ceramic to deformation at high temperatures makes possible a simple construction with only 15 components.

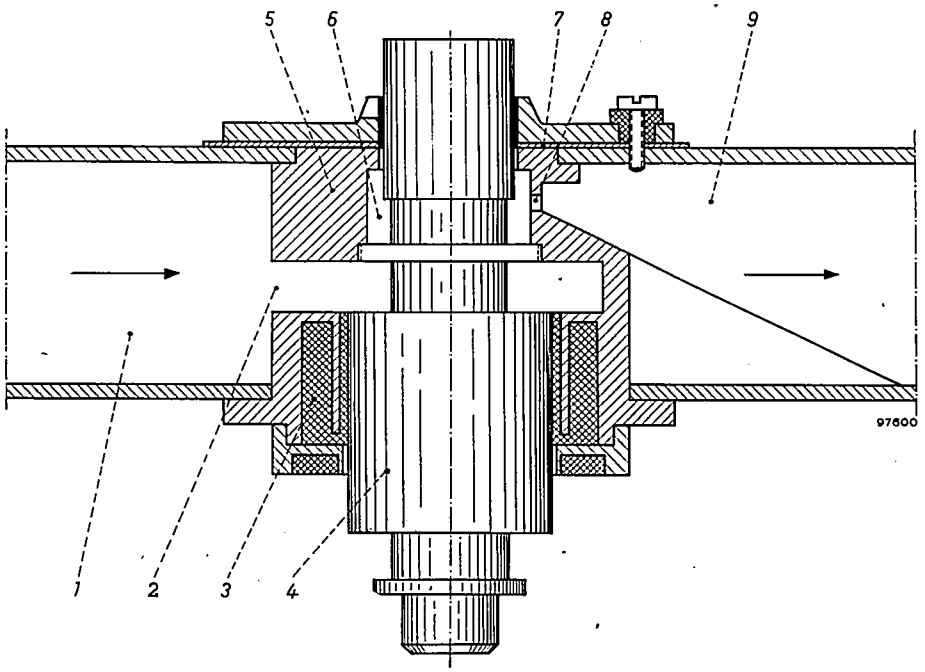


Fig. 9. Illustrating the principle of an amplifier based on the OZ 92 incorporated in a wave-guide system. *1* amplifier input. *2* input circuit. *3* "choke" insulation. *4* disc-seal triode OZ 92. *5* amplifier block. *6* output resonant cavity. *7* decoupling capacitor. *8* coupling slot. *9* output waveguide.

Summary. Description of an experimental disc-seal triode developed in the Valvo laboratory in Hamburg and provisionally designated as type OZ 92. Its electrode system is the same as in the EC 59, but its wall is of ceramic (Al_2O_3) and metal (an Fe-Ni-Co alloy). The advantages of ceramic over glass are: 1) higher thermal loading permissible owing to better heat conduction, and no electrolysis, even at high temperatures; 2) lower dielectric losses; 3) greater resistance to deformation at high temperatures. These advantages have led to a valve of very simple design, capable of delivering more than 10 W at 5000 Mc/s. The anode dissipation is 125 W, for which either water or air cooling is adequate. Anode voltage 500 V. The maximum permissible temperature of the metal-ceramic seals is 250 °C.

VOLUME 21, 1959/60, No. 4/5 pp. 109-156 Published 12th March 1960

Philips Technical Review

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

INSTRUMENTATION FOR A SUBCRITICAL HOMOGENEOUS SUSPENSION REACTOR

I. REASONS BEHIND THE CHOICE OF A HOMOGENEOUS SUSPENSION REACTOR
II. MEASUREMENT AND CONTROL OF OPERATING PARAMETERS.
IIIA. THE MONITORING OF LOW NEUTRON FLUX BY MEANS OF FAST PULSE-COUNTING CHANNELS
IIIB. THE MONITORING OF HIGH NEUTRON FLUX WITH THE AID OF AN ELECTROMETER
IV. THE SAFETY CIRCUITS

Under the direction of Prof. J. J. Went, in the laboratories of N.V. Keuring van Electrotechnische Materialen ("KEMA"), Arnhem, an unusual type of nuclear reactor is being developed, viz. a homogeneous suspension reactor. Comprehensive measurements have been made on a small subcritical model (which has operated close to the critical state). The results led to the decision to construct a second experimental model, on which various chemical processes can be studied and which can generate a power of several hundred kW.

The development work, commissioned by the Netherlands Reactor Centre, is being carried out by a team composed mainly of research staff from the KEMA and the Reactor Centre. The instrumentation for the reactor is the fruit of close cooperation between this team and Philips, and was built partly by KEMA and partly in the Philips Research Laboratories, Eindhoven.

In the first article of the series printed below, Dr. Went explains the background to this research project. The second article deals with the measurement of non-nuclear-physical operating parameters on the subcritical reactor. The other articles are concerned with the instruments designed for neutron-flux measurements and for safety purposes.

I. REASONS BEHIND THE CHOICE OF A HOMOGENEOUS SUSPENSION REACTOR

by J. J. WENT *).

621.039.52:621.039.54

The conversion of nuclear energy into energy of a useful form takes place in a nuclear reactor. The importance of this new type of energy source, in view of the world's dwindling reserves of conventional fuels, needs no emphasis here. Although various high-power reactors are in course of construction at the present time, it is by no means established which of the different fundamental possibilities is to be preferred (not to mention the purely constructional details). Investigation of these fundamental matters is still in full progress. Among the countries so occupied, The Netherlands is adding her contribution. The *credo* behind the unconventional type of reactor being developed by the KEMA — the homogeneous suspension reactor — and the reasons why initially a subcritical model was built (i.e. one that delivers no energy), will be set forth in this article.

The basic principles of nuclear reactors are to be found in the many books and articles on the sub-

*) N.V. KEMA, Arnhem, The Netherlands.

ject [1]), and will not be discussed here. Very generally the following are some of the significant points that need special attention in the development of a nuclear reactor.

Reactor physics:

1) the conditions under which the reactor becomes critical (i.e. under which a fission chain reaction is sustained);
2) the kinetic behaviour of the reactor;
3) the degree to which fresh fissile material can be "bred" during operation.

Engineering, chemical and materials problems:

1) the efficiency and the simplicity of the means of extracting heat from the reactor;
2) the form in which the fissile material is present in the reactor (in a "fuel element" or in a circulating liquid);
3) keeping the radioactive fission products safely together;
4) the simplicity with which the fission products can be removed from the reactor;
5) corrosion problems in the reactor.

Review of reactor systems

To explain the choice of a homogeneous suspension type of reactor, it will be helpful to review and classify the various reactor systems. Although any classification is bound to be subjective, the following will be useful for our purpose.

"Fast" reactors and "thermal" reactors

The first question to be decided is whether the reactor is to be a "fast" or a "thermal" type, that is whether the nuclear fissions are to be effected by "fast" or by "thermal" (slow) neutrons. Neutrons are said to be *fast* when their kinetic energy is about 2 MeV and therefore equal to the energy acquired by neutrons released in the fission process. When fast neutrons are slowed-down in a moderator until their speed ("temperature") corresponds to the temperature of the moderator — i.e. about 0.025 eV at 300 °K — they are called *thermal* neutrons. As regards the probability of undergoing fission by capturing a fast or a thermal neutron, the isotopes of uranium, plutonium (Pu) and thorium behave very differently. This probability is expressed in the *effective fission cross-section* σ_f, the unit for which is the *barn* (= 10^{-24} cm^2). In *Table I* the values of σ_f are given for some U, Pu and Th isotopes in the case of fast and thermal neutrons. The large cross-section which ^{235}U, ^{239}Pu and ^{233}U possess for thermal neutrons implies that a thermal reactor is much easier to bring into the critical state than a fast reactor. On the other hand a fast reactor has the advantage that, when a fissile nucleus captures a fast neutron, the average number of fresh neutrons released is greater. This average number, η, is also given in Table I. A large η favours the breeding of new fissile material, as will presently be explained.

Table I. Effective fission cross-section σ_f of various fissile nuclei for thermal and for fast neutrons, and the average number η of neutrons released per captured neutron.

Nucleus	Thermal neutrons of 0.025 eV (300 °K)		Fast neutrons of 2 MeV	
	σ_f barns	η	σ_f barns	η
^{235}U	580	2.08	1.3	2.3
^{238}U	0	—	0.5	2.55
^{239}Pu	750	2.06	2.0	2.8
^{232}Th	0	—	0.1	—
^{233}U	530	2.28	2.1	2.4

Thermal reactors of the heterogeneous and homogeneous types, with solid or liquid fuel

Thermal reactors can be subdivided into two groups in two ways. The first takes account of the manner in which the fissile and moderator materials are distributed, i.e. either heterogeneously (separated from each other), or homogeneously mixed. The reactors are accordingly referred to as heterogeneous or homogeneous thermal reactors, as the case may be. Another classification is concerned with the state of aggregation of the fissile material, i.e. whether it is in solid or liquid form. Although these two classifications need not run parallel, the following two combinations are nevertheless the ones that first come to mind:

1) the heterogeneous thermal reactor with solid fuel in the form of fuel elements, surrounded by the moderator, or
2) the homogeneous thermal reactor with liquid fuel, i.e. with the fissile material dissolved or suspended in e.g. water (H_2O or D_2O), which acts as the moderator.

A third combination, however, is also practicable, and was in fact under investigation, namely a heterogeneous system using a liquid fuel; for example a graphite-moderated system through which a solution of ^{233}U in molten bismuth is circulated (known as the "liquid-metal fuel reactor").

[1]) See e.g. D. J. Littler and J. F. Raffle, An introduction to reactor physics, Pergamon Press, London 1957; S. Glasstone, Source book on atomic energy, Macmillan, London 1958. For a very abbreviated introduction, see Philips tech. Rev. **19**, 246-248 and 256-257, 1957/58.

The "breeding" of fresh fissile material; conversion factor

Before going into the pros and cons of the groups into which we have just classified nuclear reactors, we shall briefly discuss the fissile materials used. This will clarify the classification of reactors into fast and thermal types, and in particular will explain why our work has been concentrated on the homogeneous thermal reactor using liquid fuel.

The fissile isotopes that occur naturally are ^{238}U and ^{235}U (found in natural uranium in the percentages 99.3 and 0.7%, respectively) and ^{232}Th, the latter in a quantity comparable to that of natural uranium. Of these only ^{235}U is fissionable both by thermal and fast neutrons (see Table I). ^{238}U and ^{232}Th can be split only by fast neutrons, and even then the probability of this happening is extremely small. This does not mean, however, that ^{238}U and ^{232}Th are of no use in a thermal reactor. Neutrons can be captured by nuclei of ^{238}U and ^{232}Th, giving rise to unstable isotopes. These decay via a series of disintegrations, involving β^- emissions, into fresh fissile isotopes, ^{239}Pu and ^{233}U, respectively:

$$^{238}_{92}U + {}^{1}_{0}n \rightarrow {}^{239}_{92}U \xrightarrow[23.5\ \text{min}]{\beta^-} {}^{239}_{93}Np \xrightarrow[2.33\ \text{days}]{\beta^-} {}^{239}_{94}Pu,$$

$$^{232}_{90}Th + {}^{1}_{0}n \rightarrow {}^{233}_{90}Th \xrightarrow[23.5\ \text{min}]{\beta^-} {}^{233}_{91}Pa \xrightarrow[27.4\ \text{days}]{\beta^-} {}^{233}_{92}U$$

(n = neutron, Np = neptunium, Pa = protactinium; the times mentioned are the half-lives for the β disintegration). Thus, fresh fissile material is "bred" in the reactor, which is accordingly called a "breeder". These breeding reactions are important in the first place in that a greater fraction of the invested fissile material undergoes fission. Another important point is that the fresh fissile materials can be bred in sufficient quantities to start other nuclear reactors after the needs of the production reactor have been met.

The efficiency of the breeding process is expressed by the *conversion ratio C*, which is the ratio of the average number of fissile nuclei of ^{239}Pu or ^{233}U formed per second from ^{238}U or ^{232}Th, to the average number of fissile nuclei ^{235}U, ^{239}Pu or ^{233}U destroyed per second. For example, $C = 0.80$ means that the reactor, by breeding, meets 80% of its own fuel requirements, so that in principle, apart from the readily available quantity of "fertile" material, i.e. ^{238}U or ^{232}Th, the reactor needs to be supplied with only 20% of the amount of fissile material that would otherwise be necessary. $C = 1$ means that the breeding process completely meets the reactor's fuel requirements, and $C > 1$ means that the reactor produces an excess of fissile material.

Since fissile materials are much dearer than fertile or source material, the general tendency will be to economize on fissile material as far as possible by adopting the breeding principle, that is to build reactors having a high conversion ratio. There are economic limits to this, however. In the case of a reactor for a power station, for example, there is no object in cutting down on fuel costs unless, per kWh produced, they represent a considerable percentage of the total cost of a kWh. The limit will approximately be reached at $C = 0.95$ (where the invested fissile material constitutes only $^1/_{20}$th of the amount consumed). There is accordingly little point in trying to achieve a higher value of C, unless it is possible to make the reactor produce so much excess fissile material as to make it worth while employing this excess in other reactors.

The value of C obtainable is closely related to the magnitude of η (see table I). Of the average number of neutrons, η, released when a fissile nucleus captures a neutron, one serves in the first place for being captured in its turn by another fissile nucleus, thus sustaining the chain reaction. Of the remainder, $\eta - 1$, part is lost through various causes, such as "leakage" at the surface of the reactor proper and absorption in the moderator, in the structural materials and possibly in the fission products, and part is available for the conversion of ^{238}U into ^{239}Pu or of ^{232}Th into ^{233}U. Suppose that $C = 1$; then for compensating the leakage and absorption losses per fission there remain only $\eta - 2$ neutrons available, and these will be sufficient only if η is materially larger than 2. Referring to the values of η given in Table I, we find the following.

1) For fast neutrons η is always greater than for slow.
2) For fast neutrons ^{239}Pu has by far the largest η, namely 2.8.
3) For thermal neutrons ^{233}U has by far the largest η, namely 2.28.
4) For thermal neutrons ^{239}Pu has a small η (i.e. 2.06) and — which is not to be seen from the table — this value drops rapidly as the neutron energy increases. At an energy corresponding to 600 °K, for example, η has already dropped considerably below 2.

The above figures lead to several important conclusions, which are summarized below.

a) A *fast* reactor fuelled by ^{239}Pu, bred from ^{238}U, can have a conversion ratio substantially higher than 1. Because of the small σ_f of ^{239}Pu for fast

neutrons, however, the density of ^{239}Pu atoms in the reactor will have to be very great to achieve criticality (*fig. 1*). Highly enriched plutonium must therefore be used. The fertile material (in this case ^{238}U) surrounds the Pu as a blanket, in which leaking neutrons are usefully employed. From time to time the generated Pu must be chemically separated from the blanket and suitably processed before being supplied to the reactor as fuel.

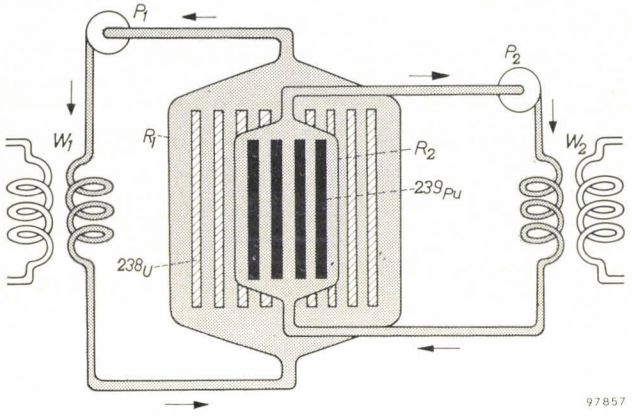


Fig. 1. Schematic cross-section of a fast reactor fuelled by ^{239}Pu bred in the reactor from ^{238}U. R_1 outer reactor vessel containing elements of fertile material (solid ^{238}U). R_2 inner reactor vessel (core) with fuel elements containing solid ^{239}Pu. Liquid sodium is circulated as coolant through both vessels. P_1, P_2 circulation pumps. W_1, W_2 heat exchangers, delivering steam.

b) A *thermal* reactor fuelled by ^{233}U, bred from ^{232}Th, can have a conversion ratio of $C = 1$ provided the neutron conservation is adequate. Owing to the large effective cross-section of ^{233}U for fission by thermal neutrons ($\sigma_f = 530$, see Table I), it is possible to make this kind of reactor critical even with quite low concentrations of ^{233}U. Thus the fertile material can also be accommodated inside the reactor system, so that the fuel material bred from it is immediately available for fission. This dispenses with the need for chemical separation and the supply of fresh fissile material, which is unavoidable under (*a*).

c) Although we have specially emphasized under (*a*) and (*b*) the importance of the fissile materials ^{239}Pu and ^{233}U, which can be bred in fast and thermal reactors, it should be remembered that ^{235}U is likely to remain the basic fuel of nuclear reactors for some time to come. This is after all the only directly fissile material that occurs naturally, and for this reason constitutes at present the basic material for any nuclear energy programme.

Classification of reactors according to conversion ratio

Since the conversion ratio C is one of the quantities determining the choice of a given type of nuclear reactor, it will also be useful to classify reactors according to the magnitude of C.

1) Reactors operating exclusively with highly enriched fissile materials. This type will be preferred where small, mobile reactors are required and where the price of the fissile material is not a primary consideration. Such reactors employ scarcely any fertile material, and therefore C will be zero or very small.
2) Normal thermal reactors, fuelled by ^{235}U or by ^{239}Pu bred from ^{238}U, in which a conversion ratio between 0.6 and 0.85 is often obtainable. Such a value is clearly favourable for fuel economy.
3) Thermal reactors breeding ^{233}U from ^{232}Th and with the utmost in neutron economy, having a value of C from 0.9 to 1.1. In these reactors the fraction of the original charge of fissile material that undergoes fission is high enough to make the use of thorium practicable. Provided the chemical processing involved is not excessively costly, the fuel costs can in fact be very low.
4) Fast reactors that breed ^{239}Pu from ^{238}U (fig. 1), with C values from 1.5 up to 1.8. Such reactors are capable of producing so much ^{239}Pu as to enable the excess to be employed for fuelling other reactors, fast or thermal. The practical value of this is somewhat limited, however, by the fact that up till now it seems difficult to use Pu in thermal reactors, for one reason because of the unfavourable value of η (see above).

Choice of reactor type

An important quantity in a nuclear reactor is the *multiplication factor* k of the neutrons, which is the ratio of the number of neutrons in a given "generation" to the number in the preceding generation, the lifetime of a generation being about 10^{-3} sec in thermal reactors and about 10^{-7} sec in fast reactors. The value $k = 1$ is required if the chain reaction is to be just self-sustaining. If $k < 1$, the number of neutrons present decreases with time. If $k > 1$, the number increases and the reactor is said to have a positive *reactivity*. The latter quantity is defined as $(k - 1)/k$.

When changes in reactivity occur, no matter what the cause, the kinetic behaviour of the reactor becomes of importance; for certain fundamental reasons this behaviour can give rise to greater difficulties with fast than with thermal reactors. Reactor kinetics are closely bound up with safety. In a densely populated country like The Nether-

lands, safety is a particularly serious consideration, and our attention was therefore turned primarily to thermal reactors. On the other hand, greater economy in the use of reactor fuel is certainly not devoid of interest, especially in The Netherlands which is poor in fissile material. The obvious course, then, was to consider a thermal type of reactor, capable of breeding ^{233}U as fuel from ^{232}Th whilst making the most economical use of neutrons. For starting this reactor, ^{235}U or ^{239}Pu might be used.

The moderator would have to be chosen for optimum neutron economy. The best in this respect is heavy water. Furthermore, as little use as possible should be made of neutron-absorbent structural materials inside the reactor. The absorption of neutrons in the fission products and the leakage of neutrons from the reactor should be minimized.

All these requirements point towards a homogeneous reactor with the fissile material distributed in heavy water. In the core of this type of reactor there is no need for any structural materials. For the purpose of heat extraction the homogeneous reactor medium circulates through an external heat-exchanger. The direct transfer of energy from the fluid medium can be highly efficient, being proportional to the temperature difference (which is limited, of course) between the inlet and outlet, and to the pumping speed, which is subject to merely practical restrictions. A power extraction of 40 MW per cubic metre reactor volume seems well within the bounds of possibility.

Advantages can further be derived from the circulatory system by passing the reactor medium through a decontamination plant to remove the strongly neutron-absorbent fission products, such as ^{135}Xe and ^{149}Sm, so rapidly that only few neutrons are lost to these substances. In heterogeneous reactors the fission products are retained in the fuel elements and the neutron losses are consequently higher.

Comparison of homogeneous and heterogeneous reactors

Technologically, a homogeneous reactor of the type described has little in common with the more conventional heterogeneous types. The main difference is that it does not contain the costly fuel elements, with their limited lifetime, that have to be installed in heterogeneous reactors. On the other hand, these fuel elements have the merit of retaining the radioactive fission products, which in a homogeneous reactor are freely circulated. Since part of the circulatory system is outside the reactor proper, namely the heat exchanger and the circulation pump, the system must be completely leak-tight, to prevent the dangerous dispersion of radioactive substances; the demands in this respect are even higher than those made on the protective shield around the fuel elements in a heterogeneous reactor, which are entirely enclosed in the core.

Another notable difference concerns the control and safety systems. To keep the neutron flux constant — and hence the generated power — *heterogeneous* reactors are equipped with a control mechanism consisting of one or more rods of a neutron-absorbent material, such as boron. The neutron flux is controlled by varying the depth to which these rods are inserted in the core. A safety mechanism is required for automatically shutting-down a heterogeneous reactor if the neutron flux should become excessively high. This may be done by causing the control rods, or similar "safety rods", to drop into the reactor [2]), or by "poisoning" the reactor with a boron injection. An important advantage of the *homogeneous* type of reactor described is that it does not require control and safety rods. There are two reasons for this. The nuclear energy is released primarily as kinetic energy from the nuclear fragments formed upon fission, and this energy in a homogeneous reactor is directly transferred to the water. Now the first favourable circumstance here is the high thermal expansion coefficient of water; the expansion reduces the moderating action of the water so as to give the reactivity a high negative temperature coefficient. This is also the case, of course, in water-moderated reactors of the heterogeneous type. Secondly — and this applies solely to homogeneous reactors — the temperature coefficient works promptly, owing to the virtually immediate transfer of energy from the fission fragments to the water. The consequence is that the production of neutrons is automatically retarded by the rising temperature.

In most of their other technical features, too, homogeneous and heterogeneous reactors are completely disparate. *Table II* gives a schematic comparison of both types in terms of complexity. Without allocating a specific weighting factor to each of the points mentioned, and thereby evaluating some numerical measure of the degree of simplicity, it cannot be said which type is to be preferred. A proper assessment must necessarily be based on experience of the construction and operation of both types, which should preferably be at least as large as a power demonstration reactor. Heterogeneous reactors of this size are already fairly

[2]) See e.g. M. van Tol, Monitoring, control and safety equipment for a nuclear reactor of the swimming-pool type, I. General description, and F. E. L. ten Haaf, G. Klein and F. J. Schijff, II. Further description of certain component units, Philips tech. Rev. **19**, 245-257 and 273-285, 1957/58.

Table II. Comparison of some major problems of homogeneous and heterogeneous thermal reactors, in terms of complexity.

Nature of problem	Homogeneous reactor with circulating fuel	Heterogeneous reactor with stationary fuel
Core construction	Simple	Complex
Heat extraction	Complex (leak-tight circulatory system)	Simple
Control of reactivity	Simple (negative temperature coefficient)	Complex (control and safety rods)
Control of fission-product dispersion	Complex (fission products inside and outside reactor)	Simple (fission products retained in fuel elements)
Fuel element	Simple (solution or suspension)	Complex (canned elements)
Plant maintenance	Complex (entire circulatory system is radioactive)	Simple (intense radioactivity confined to core)
Chemical processing and reprocessing	Simple	Complex
Heat extraction per m^3 volume	Very high	Moderate to high
Corrosion	Difficult problem	Difficult problem

numerous, but only very few homogeneous reactors have so far been built. It is therefore not yet possible to pronounce definite judgement on which type is more desirable for a particular application. The construction of the homogeneous subcritical reactor to be described below, in which the fuel is kept in circulation, represents an attempt to fill some of the gaps in our knowledge of homogeneous reactors.

Further details of the homogeneous reactor with circulating fuel

The concept "homogeneous reactor with circulating fuel" is still not fully defined even if we add the information that the reactor breeds ^{233}U from ^{232}Th and is moderated by heavy water to achieve maximum neutron economy. The two principal questions still to be decided are:

1) whether the reactor is to consist of two zones, i.e. of a core surrounded by a fertile (breeder) blanket, or of one zone only;
2) whether use is to be made of a solution or a suspension (slurry).

These two questions are to some extent related. Let us take question (1) first.

A reactor having a fertile blanket can in principle have a higher conversion ratio than a one-zone reactor. This is not only because the blanket turns the leaking neutrons to a useful purpose again, but also has to do with the properties of the intermediate product ^{233}Pa which is obtained in the formation of ^{233}U from ^{232}Th:

$$\underset{\sigma_c = 7.5}{^{232}_{90}\mathrm{Th}} + {}^{1}_{0}\mathrm{n} \rightarrow \underset{1400}{^{233}_{90}\mathrm{Th}} \xrightarrow[23.5\ \mathrm{min}]{\beta^-} \underset{70}{^{233}_{91}\mathrm{Pa}} \xrightarrow[27.4\ \mathrm{days}]{\beta^-} \underset{55\ \mathrm{barns}}{^{233}_{92}\mathrm{U}}.$$

The neutron-capture cross-section, σ_c, of ^{233}Pa is 70 barns as against 7.5 barns for ^{232}Th, and the disintegration of ^{233}Pa into ^{233}U has a half-life of almost a month. During this relatively long half-life there is thus an appreciable probability that the ^{233}Pa nuclei will capture neutrons, producing ^{234}Pa, which does *not* transform into the fissile ^{233}U (fissile because its $\sigma_f = 530$ barns $\gg \sigma_c$). The probability of the unwanted formation of ^{234}Pa depends not only on the σ_c of ^{233}Pa and its half-life but also on the neutron flux at the point considered, to which the probability is proportional. If steps are taken to ensure that the ^{233}Pa (and hence also the fertile material ^{232}Th) only occurs where the neutron flux is low, i.e. on the outside of the reactor, the loss of ^{233}Pa, and of neutrons for the chain reaction, resulting from the formation of ^{234}Pa, will be small. This consideration would therefore lead to the arrangement of the fertile ^{232}Th in the form of a blanket around a core of the fissile ^{233}U, i.e. to a reactor composed of two zones (*fig. 2a*).

The reactor in course of development in The Netherlands is *not* being built on this principle. This is mainly because there are two serious drawbacks attaching to reactors with two zones: their construction is more complicated than that of one-zone reactors (compare fig. 2*a* with fig. 2*b*), and to find a suitable material for the core vessel is a most difficult problem (see below). The advantage set against these drawbacks, i.e. the higher conversion ratio, is not of great significance in our case, since it would result in only a relatively slight reduction of the cost of a kWh (the cost of the fissile material being a relatively small percentage of the total).

The structural complexity of two-zone reactors

arises from the necessity of providing them with two distinct cooling circuits and with two different chemical processing systems (see fig. 2*a*). The latter are needed for the following reasons.

a) Chemically pure ^{233}U must be separated from the blanket and then supplied, in the appropriate form, to the reactor core in order to replenish the spent fuel. At the same time the blanket must be rid of the products of corrosion, erosion and fission (erosion is caused by the solid suspended particles when a suspension is used). Apart from being an unwanted complication, the chemical processing which this involves adversely affects the conversion ratio, since it is inevitably accompanied by some loss of ^{233}U and thus the "effective" conversion ratio will be somewhat lower than the "physical" conversion ratio.

b) The reactor core must be rid of fission products such as ^{135}Xe, which would be harmful to the conversion, and also of corrosion and erosion products.

On the other hand, in a one-zone reactor (fig. 2*b*) the generated ^{233}U can participate directly in the fission process, without any intermediate stage. The chemical separation plant referred to under (*a*) is therefore not required, which represents a considerable simplification.

The dividing wall between the core and the blanket in a two-zone reactor must obviously be highly permeable to neutrons. It must also be resistant to corrosion from various sources, namely from the liquids with which it is in contact, high temperatures, β and γ radiation, high neutron flux and bombardment by fission products. As yet the best compromise between permeability to neutrons and resistance to corrosion is a zirconium alloy known as zircaloy. It is doubtful, however, whether this material would prove sufficiently resistant to corrosion under the conditions mentioned.

It is a simpler matter to find a good material for the wall of a one-zone reactor, the requirement of permeability to neutrons not being applicable here. The wall can be protected against fast neutrons and high-energy fission products by interposing a "thermal shield" of a material that slows neutrons down to low velocities.

We now come to the question as to the form in which the fissile material should be circulated: as a solution or as a suspension. Let us first consider the two-zone reactor. In its *core* this type requires fissile material only, and that in a very low concentration (about 0.05% by volume), fertile material being undesirable there in view of the ^{234}Pa that would then be formed. If we choose the fissile material in the form of oxide particles suspended in water, it is difficult, owing to the high specific gravity of these particles (approx. 10 g/cm^3), to maintain with such a low concentration a sufficiently homogeneous distribution throughout the primary circulatory system. With a solution in water of, say, uranyl sulphate (UO_2SO_4) it would be easier to maintain the necessary homogeneity, provided at least the temperature and the radiation caused no precipitation of the solution. In the *blanket*, on the other hand, the concentration of fertile material must be high (5 to 10% by volume), in order to absorb all the neutrons leaking from the core. The relevant Th or U salts, however, are not sufficiently soluble, and moreover concentrated sulphate solutions would be excessively corrosive. This means that only a suspension (of oxide particles) can be used in the blanket of a two-zone reactor.

The above also applies to a one-zone reactor in which the fissile and fertile materials are mixed (approx. concentration 3% by volume).

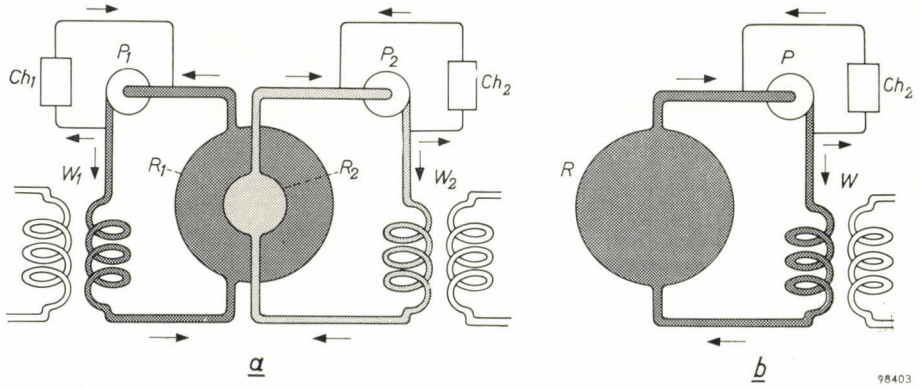


Fig. 2. Schematic representation of two homogeneous reactors.
a) Two-zone homogeneous reactor: outer vessel R_1 with a circulating suspension of ThO_2 (breeding blanket), inner vessel R_2 with a circulating solution of UO_2SO_4 in water. P_1, P_2 circulation pumps. W_1, W_2 heat exchangers. Ch_1 chemical plant which 1) separates ^{233}U and ^{232}Th and processes the ^{233}U, and 2) removes products of corrosion, erosion and fission. Ch_2 chemical plant for removing fission, corrosion and erosion products.
b) One-zone reactor, with vessel R, circulation pump P and heat exchanger W through which circulates a suspension of ThO_2-UO_2 particles (fertile and fissile material) in water. Ch_2 chemical plant as in (*a*). A great advantage of the one-zone reactor is that it does not require the chemical plant Ch_1 as in (*a*), nor does it involve the difficulty of finding a suitable material for the wall of R_2.

The question, "solution or suspension?", can thus be answered as follows: a two-zone reactor can probably operate on a solution and a suspension; a one-zone reactor that also produces fresh fissile material must operate on a suspension.

For special purposes one can also have a one-zone reactor that does not produce fresh fissile material, i.e. one that only generates heat, without conversion necessarily taking place. This type, called a "burner" reactor, can operate on pure ^{235}U or ^{239}Pu. By virtue of the arguments given above for the core of a two-zone reactor, a very dilute solution can be used in this case.

To give the reader an idea of the possible dimensions and of the invested quantities of nuclear and moderator (D_2O) materials, we have set out in *Table III* some general data concerning a burner reactor, a two-zone breeder reactor and a one-zone breeder reactor. The figures assume a maximum heat production of 440 MW at a maximum temperature of 300 °C, giving an electric power output of about 100 MW. The amounts of nuclear and moderator material take into account the required cooling circuit outside the reactor vessel. The table reveals an advantage of the two-zone reactor not hitherto mentioned, namely that this type calls for a substantially smaller investment in nuclear material and heavy water than the one-zone type. At present this advantage does not offset the drawbacks mentioned above, no more than does the higher conversion ratio.

Table III. Broad comparison of three types of homogeneous reactor with circulating fuel. Heat production 440 MW at 300 °C max.

Reactor type / Feature	Burner reactor with ^{235}U or ^{239}Pu	Breeder reactors with ^{232}Th		
	1 zone	2 zones		1 zone
		core	blanket	
Fissile material	$^{235}UO_2SO_4$ solution	UO_2SO_4		ThO_2-UO_2
Fertile material	—		ThO_2	
Concentration (grams/litre)	2	6	500-1000	300
Concentration of fissile material (%)	100	100	very low	1.5
Diameter of vessel (metres)	3	1.5	3	3.6
Conversion ratio C	0	approx. 1.1		approx. 1.0
Quantity of D_2O (metric tons)	30	30		45
Quantity of fissile material (kg)	65	90		200
Quantity of fertile material (metric tons)	0	11		14

The design and development of the homogeneous one-zone suspension reactor

Having been led by the above considerations to decide in favour of a homogeneous one-zone suspension-type reactor, we were faced by problems that may be divided into three groups:

1) Problems concerning the stability of the reactor (also during start-up and shut-down).
2) Problems relating to the chemical processes in the reactor and the separation of impurities in the ancillary circuits, including the removal of corrosion products.
3) General engineering and structural problems.

With a view to solving the problems under (1) a small subcritical reactor was built. This will henceforth be called reactor I, and will frequently be referred to in the subsequent articles in this series. The results of the tests on reactor I were encouraging and a second, smaller reactor is now under construction (reactor II), designed for a power output of about 250 kW. This will be used for studying the processes mentioned under (2). The third step, it is hoped, will be the construction of a prototype of a high-power one-zone suspension reactor.

Reactor I operates on the principle illustrated in fig. 2*b*. The system is shown again in *fig. 3*, with some important details added. By means of the valve T_1 a variable amount of the circulated suspension can be diverted through the shunt pipe *Sh*, thus regulating the flow through the reactor vessel *R* (the circulation pump *P* operates at a constant speed). Fig. 3 also shows how the concentration of fissile material in the suspension is controlled. To withdraw fissile material from the suspension,

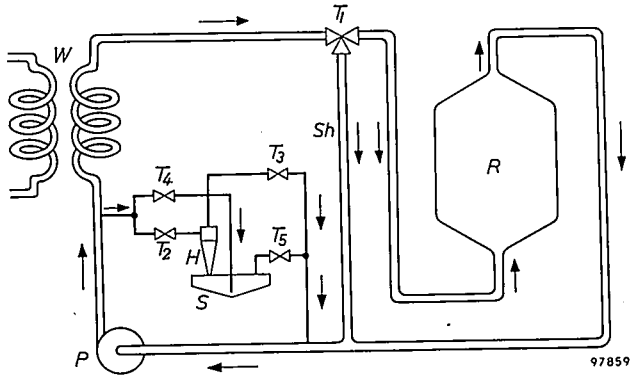


Fig. 3. Simplified diagram of the subcritical suspension reactor (reactor I) built by KEMA. *R* reactor vessel. *P* circulation pump. *W* two-stage heat exchanger (cooler) which transfers its heat via a closed water circuit to the main cooling water. *Sh* shunt pipe across reactor vessel. T_1 control valve. *H* hydrocyclone which, when valves T_2 and T_3 are open, separates fuel particles from the suspension. *S* reservoir in which the separated fuel is collected. To add fissile material to the reactor circuit it is necessary to open valves T_4 and T_5.

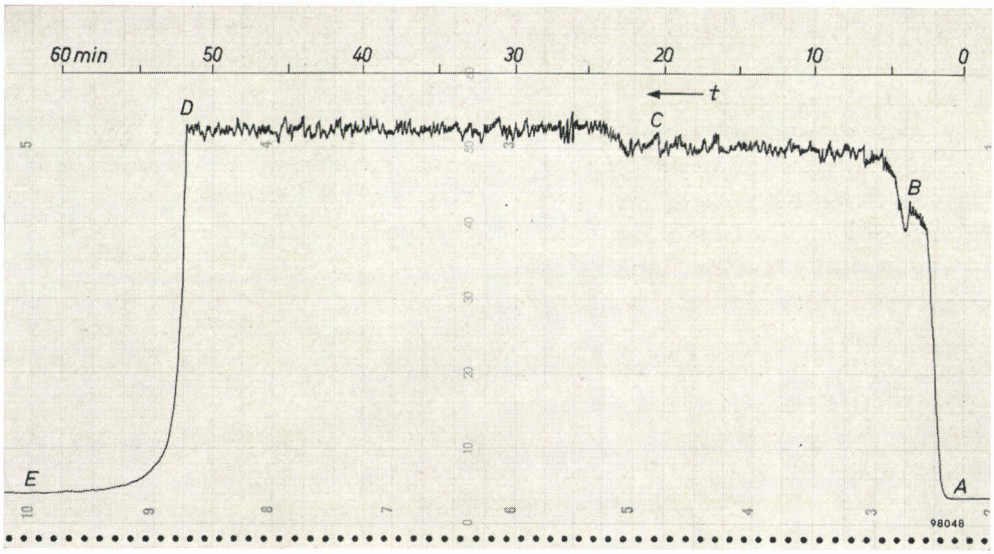


Fig. 4. Recording of the neutron flux Φ in the KEMA subcritical reactor, as a function of time t. To the right of A the neutron flux had the initial value Φ_0, delivered by an external neutron source; at this point the reactor contained water only.

A marks the point at which UO_2 particles were introduced into the reactor circuit from the reservoir; Φ rapidly increased as a result of the fission of ^{235}U nuclei. At B the connection with the reservoir was momentarily interrupted, at C closed. At D the hydrocyclone came into operation, removing fuel particles from the suspension. After some minutes, at E, the reactor vessel was again depleted of fuel, returning the neutron flux to the starting value Φ_0. The fluctuations in the curve are due to statistical fluctuations in Φ.

valves T_2 and T_3 are opened. Part of the suspension then flows via T_2 to a hydrocyclone H in which the solid particles are centrifugally separated from the water. These fuel particles are collected in a reservoir S, whilst the water returns via T_3 into the main circuit. To draw fresh fuel into circulation from the reservoir, valves T_4 and T_5 are opened; a jet of water then enters the reservoir and forces fuel particles into the main circuit. In this way the fuel concentration can be very accurately controlled.

Fig. 4 shows a recording of the neutron flux Φ in reactor I. D marks the point where the valves T_2 and T_3 (fig. 3) were opened, setting the hydrocyclone into operation and causing Φ to drop rapidly to the initial value Φ_0, supplied by an external source of neutrons (see below).

We shall now deal at greater length with the stability problem, mentioned above, after which we shall discuss briefly the chemical, materials and engineering questions.

Physical stability

We have seen that, owing to the high negative temperature coefficient of its reactivity and the promptness with which this becomes effective, a water-moderated homogeneous reactor is inherently stable. The reactivity is regulated by corresponding changes in the temperature of the reactor. Consequently, no control and safety mechanism is required, a fundamental physical mechanism — the thermal expansion of water — being responsible for control and safety in this type of reactor. There is thus no possibility of a dangerous thermal "run-away", as there is in a heterogeneous reactor if the control system is defective.

What has, however, to be taken into account is the danger that a sudden transition to a higher equilibrium temperature resulting from a sudden increase in the concentration of the fissile material might cause a wave of high pressure that could damage the reactor vessel. It is therefore necessary to determine the maximum magnitude and speed of temperature variations that can arise in such a reactor, both under normal and abnormal conditions. An abnormal condition would occur if, e.g., the reactor circulation were to be suddenly stopped without the fuel particles first having been withdrawn from the reactor. Even greater difficulties would result if, after such an abrupt shut-down, the reactor were to be suddenly started again. In that case the distribution of the particles in the main circuit would be initially very heterogeneous and would only gradually become homogeneous. The consequence would be temporarily severe changes in reactivity, hence in neutron flux. Investigations in this connection should consist of measuring the neutron-flux variations in a reactor which is either slightly subcritical, or critical at so low a power level as to keep the radioactivity of the reactor core as low as possible and

prevent large temperature excursions. In the first place this will mean that, owing to the extremely low power production, so little heat will be generated that the above-mentioned stabilizing action of temperature fluctuations will not be effective; the neutron-flux variations will thus be relatively greater and can therefore be more accurately measured. In the second place the low level of radioactivity will enable modifications to be made to the installation more readily.

Not only is it desirable, for the reasons mentioned, to limit the temperature excursions; as one of the principal operating parameters, the temperature must in fact be kept accurately constant. This is done by an electronic regulating device to be described in article II of this series.

In these tests, then, temperature variations with their stabilizing effect are deliberately excluded, and so some other safety mechanism must be adopted, e.g. normal control and safety rods containing boron. Three such rods are accordingly fitted in reactor I; they are mounted in the neutron reflector which surrounds the reactor core (*fig. 5*) and which will be touched on presently. The insertion of these rods can reduce the reactivity by up to 6%.

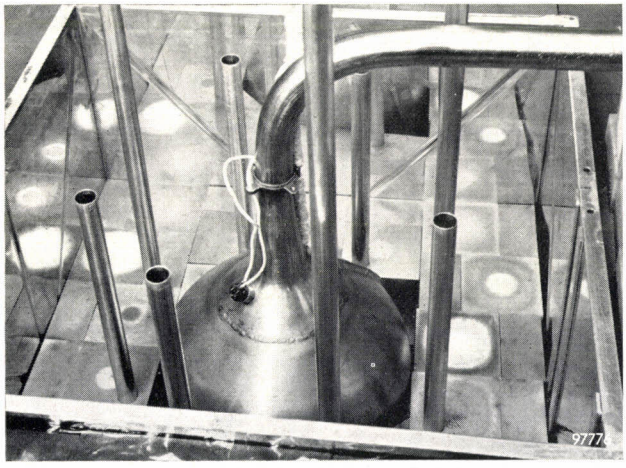


Fig. 5. The reactor vessel in the KEMA reactor I. Of the eight vertical tubes three serve as guides for the control and safety rods; the five shorter tubes each contain a neutron detector. The blocks (partly removed) inside the four walls are of beryllium oxide; together with blocks of graphite outside the walls, they constitute a neutron reflector. The reactor vessel is of stainless steel; it has a capacity of 18 litres and measures 30 cm in diameter. A temperature sensing element is fitted to it.

The neutron flux must not, of course, be made arbitrarily small, particularly because it is also required to measure transient fluctuations in the neutron flux: in a given short time a sufficient number of neutrons must have been counted to prevent statistical fluctuations in the neutron flux from influencing the measurement. We therefore stipulated that, in general, it should be possible to detect 40 000 neutrons per second.

We have referred above to measurements in a reactor in the subcritical state. This is a state in which the multiplication factor k is less than 1. The neutron flux can then be maintained only provided neutrons are continuously supplied to the system.

This can be done with the aid of a radium-beryllium source, in which the alpha rays emitted by radium react with the beryllium as follows:

$$^{4}_{2}\mathrm{He} + {}^{9}_{4}\mathrm{Be} \rightarrow {}^{12}_{6}\mathrm{C} + {}^{1}_{0}\mathrm{n}.$$

Another kind of neutron source makes use of an accelerating tube, producing the reaction

$$^{2}_{1}\mathrm{D} + {}^{2}_{1}\mathrm{D} \rightarrow {}^{3}_{2}\mathrm{He} + {}^{1}_{0}\mathrm{n}.$$

Reactor I is provided with both types of neutron source.

Suppose that the external neutron source constantly supplies N neutrons per second. After n neutron generations the total neutron production, i.e. the production of source and reactor together, will be (for $n \gg 1$):

$$Nk\,(1 + k + k^2 + \ldots + k^n) \approx N\frac{k}{1-k}$$

neutrons per second.

(Since the lifetime of one generation is approximately 10^{-3} second, this state is reached in, say, 0.1 sec.) The neutron flux caused by this production can now be measured. If k varies by an amount δk, then, where N is constant, the quantity $Nk/(1-k)$ will also vary, the more so the more k approaches unity [3]). If we measure k and also the neutron-flux variations, we can calculate δk. Of course, the temperature must be kept carefully constant during these measurements. If the temperature slowly rises — e.g. as a result of the pump energy dissipated in the water — the required quantity δk will have a variation $\delta' k$ superimposed on it, depending on the temperature increment and the negative temperature coefficient of the reactivity. For our purposes the temperature was required to remain constant within ± 0.1 °C.

Finally, a word about the determination of the conditions for criticality (the geometry of the reactor, the temperature T of the core and the concentration c of the fissile material in the water),

[3]) The effect of *delayed* neutrons is left out of account here. This is not permissible, however, if k closely approaches unity, since the delayed neutrons (about 0.7% of the total number released) apparently have a much longer lifetime (approx. 10 seconds) than the "prompt" neutrons (approx. 10^{-3} sec).

i.e. the conditions under which k becomes exactly equal to unity. Since $Nk/(1-k)$ is then infinitely large, it is a relatively simple matter to determine the required conditions. If we plot $(1-k)/N$ as a function of c, with T constant, we obtain a curve as shown in *fig. 6*. From subcritical plots we can thus find the critical concentration c_{cr} by extrapolation. The magnitude of δk for small changes in concentration is then fairly easy to derive from c_{cr}.

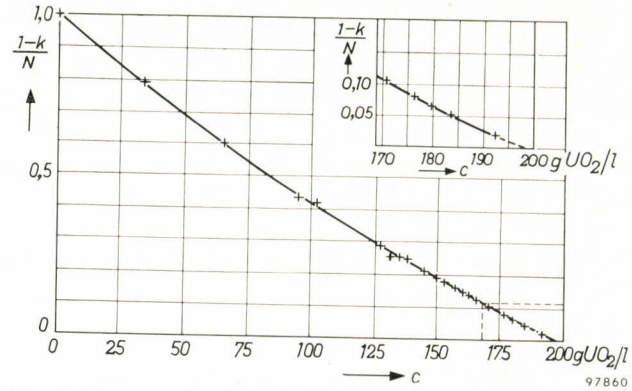


Fig. 6. The quantity $(1-k)/N$ plotted on an arbitrary scale as a function of fuel concentration c, the temperature T of the reactor medium being kept constant. The crosses represent measurements on the KEMA reactor I for a given arrangement of neutron source and neutron detector, and at $T = 25$ °C. The concentration c_{cr} at which, under these conditions, the reactor would become critical is found by extrapolation (see inset); in this case $c_{cr} = 198$ grams UO_2 (uranium with 20% ^{235}U) per litre suspension.

The following articles in this series will describe the instrumentation for the above measurements in reactor I, jointly developed by the Arnhem team and Philips Research Laboratories at Eindhoven. The electronic equipment can be seen in *fig. 7*. The following are some of the conditions which the installation was designed to meet:

a) Reactor I was to be fitted with control and safety rods.

b) The core temperature was to be kept automatically constant within $\pm$ 0.1 °C in the range from 20 to 80 °C.

c) The external neutron source had to be sufficiently intensive to allow the neutron detectors to count as many as 40 000 neutrons per second in the absence of fissile material. On the other hand it was necessary to be able to make the source so weak as not to cause $Nk/(1-k)$ to become too high, despite the high value of $k/(1-k)$.

d) The operation of the neutron detectors with their associated electronic equipment was to remain linear even at values far in excess of 40 000 neutrons per second.

e) It was necessary to be able to measure and vary continuously the concentration of the suspension.

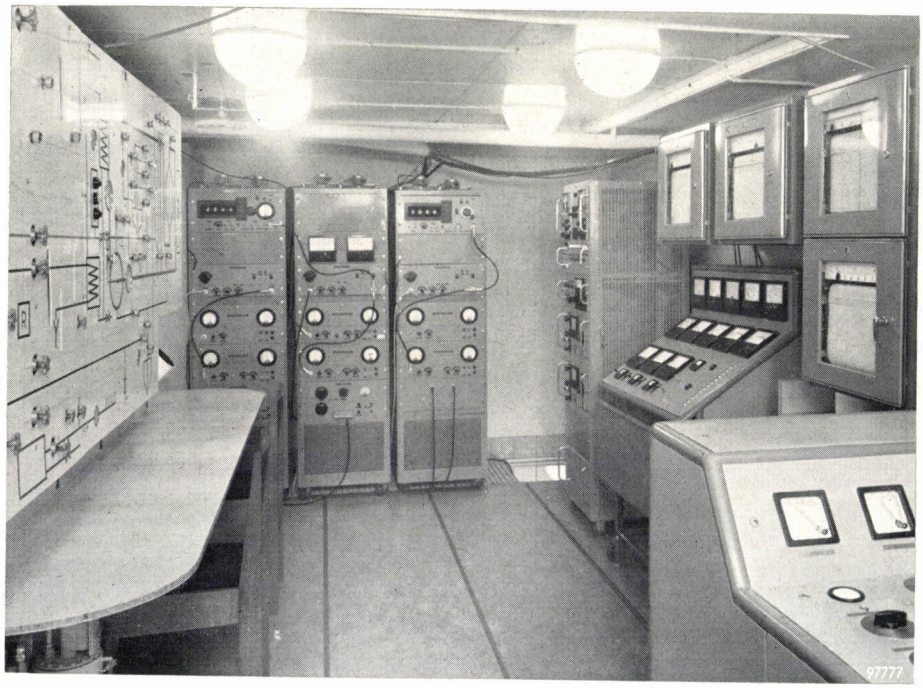


Fig. 7. Electronic part of the instrumentation for the KEMA reactor I. Left: graphic panel of installation. Centre, background: neutron counters. Right, moving towards foreground: temperature regulator, neutron-flux meters and equipment for control and safety rods; meters for recording temperature and concentration of the suspension in the reactor vessel, the rate of flow (in m^3/h), and the coolant temperature; control desk for the accelerating tube of one of the neutron sources.

Chemical processes in the reactor

Many chemical processes taking place in a suspension reactor can be studied just as well in a small as in a large reactor, provided there is not much difference in their temperature and radiation conditions, or in their suspension properties (such as concentration per unit volume and the properties of the particles). The operating temperature is set by adjusting the concentration of the suspension. Equivalent radiation conditions can be obtained by ensuring that the small reactor produces the same power per unit volume as the large one, i.e. just as many fissions per second and per cm^3. The power is adjusted by regulating the extraction of heat from the heat exchanger (fig. 2*b*). Furthermore it is desirable that the volumetric concentrations of fissile and fertile material in the small reactor should be just as high as in the large. This condition can be met if the ratio between the numbers of fissile and fertile atoms is independently adjustable.

Thus, under the conditions prevailing in a large power reactor, a small "process reactor" can be used for studying, *inter alia*: the effect of radiation on the decomposition of water, the colloid-chemical stability of the suspension and the mechanical stability of the suspended particles under the influence of the fissions that occur in them, the extraction of the fission products in processing circuits and the removal of corrosion and erosion products.

With a view to studying these processes, a second small reactor of this type (reactor II) is under construction at the KEMA laboratories. It is as small as is practicable from the viewpoint of reactivity. That there is a lower limit to the volume of the reactor vessel follows from the fact that, as the vessel is reduced in size, the surface area decreases less rapidly than the volume; neutron leakage thus becomes relatively greater and would finally become unworkably large. The lower limit of the dimensions can be reduced by compensating more or less for the increasing leakage of neutrons. To this end, the vessel in reactor I is surrounded by a neutron *reflector* of a good moderating material, consisting of a layer of beryllium oxide surrounded by graphite (fig. 5). This, in conjunction with the use of enriched fuel (20% ^{235}U), made it possible to suffice with a volume of 18 litres. With these very small dimensions less fuel is required to make the reactor critical when it is moderated with ordinary instead of heavy water. Reactor I is therefore fuelled by a suspension of uranium oxide (UO_2) in ordinary water, and so also will be reactor II, since for studying the processes in question it makes virtually no difference which kind of water is used.

The conversion ratio of reactor II cannot be high, nor is it intended to be, for the object with this reactor is to investigate quantitatively the processes that should be optimized in order to achieve a high conversion ratio in a power reactor.

General materials and engineering problems

No less important are the materials and engineering problems. In this article, however, they need only be touched on very briefly.

In the type of reactor with which we are concerned the circulating fluid is exceptionally radioactive, and moreover the material of the entire main circuit is made radioactive by delayed neutrons (otherwise left out of account here). As a result, two groups of problems are encountered that are far less serious in a reactor in which the fuel is not circulated.

First of all, steps must be taken to ensure that all components of the reactor system — pumps, valves, etc. — are absolutely and permanently leak-tight. Secondly, there is the question of maintenance. A life of 20 years is estimated for a normal electric power station. No power plant is conceivable with such a life in the absence of provisions for periodic maintenance. A nuclear reactor must therefore be designed to allow for proper maintenance — obviously with facilities for remote control.

The process reactor (II) in course of construction will provide important data on both problems. The information will not, however, be complete. Further progress must necessarily relay on experience with components on the scale of a power reactor, and on experience of the maintenance costs involved. The third step in our development programme will therefore be to build a prototype of a power reactor. It need hardly be said that this will be a very costly step indeed.

Summary. At the KEMA laboratories at Arnhem work is in progress, under the auspices of the Netherlands Reactor Centre and under the direction of the author, on the development of an unconventional type of nuclear reactor — a one-zone homogeneous suspension reactor. After distinguishing between fast and thermal reactors, and classifying the latter into heterogeneous and homogeneous types, the author discusses the "breeding" of fresh fissile materials (^{239}Pu and ^{233}U) from "fertile" materials (^{238}U and ^{232}Th, respectively). The efficiency of the breeding process is expressed in the conversion ratio, the magnitude of which is taken as a further basis for classifying reactors.

In a densely populated country like The Netherlands, with its shortage of fissile material, considerations of safety and nuclear fuel economy indicate the use of a homogeneous reactor with circulating fuel. This leaves a choice between a two-zone reactor (i.e. one with a core of fissile material surrounded by a blanket of fertile material) and a one-zone

type (in which fissile and fertile materials are mixed). The latter type was preferred, mainly because it is less complicated. The circulated fluid consists of a suspension of ThO_2 particles containing some UO_2 in heavy water, the latter also acting as moderator. The fissile material is ^{233}U, bred in the reactor itself from ^{232}Th.

The first step in the development programme was to build a small-scale reactor of this type for experiments in the subcritical state. The neutron flux is maintained by an external neutron source. The circulated fluid is a suspension of UO_2 (natural uranium with 20% ^{235}U) in ordinary water. One of the quantities determined was the fuel concentration required to achieve criticality. As a second step a similar reactor (process reactor) is being built, designed for a power output of 250 kW and on which various chemical processes will be studied under conditions comparable to those in a large power reactor. The third step will be to build a prototype power reactor of the homogeneous suspension type.

II. MEASUREMENT AND CONTROL OF OPERATING PARAMETERS

by B. L. A. van der SCHEE *) and M. van TOL.

621.039.524.46-53:621.039.524.46-79

In the experiments with the subcritical suspension reactor in one of the laboratories of N.V. KEMA at Arnhem [1]), the temperature in the reactor vessel must be kept rigorously constant (variations $<$ 0.1 °C) at any required value between room temperature and 100 °C. It is also necessary to be able to measure certain operating parameters, chief among which are, apart from the reactor-vessel temperature, the concentration of the suspension in the reactor vessel and the rate at which the suspension is pumped around the reactor circuit. In this article we shall discuss in turn the control devices for keeping the *temperature* constant [2]), the instruments for measuring the *concentration* and those for measuring the *rate of flow*. Finally we shall review the whole reactor system together with its instrumentation (but not the equipment for neutron-flux monitoring and for operating the control rods, which will be dealt with in subsequent articles).

Temperature control

The design of the temperature control system had to take account not only of the permissible maximum deviation of the temperature, but also of the fact that 3 to 7 kW of energy is supplied to the fluid by the pump that keeps the suspension circulating in the reactor circuit. This means that even at the highest operating temperature (below the boiling point) some cooling is necessary. The operating temperature is determined by the degree of cooling. The stabilizing of the temperature against variations in the power dissipated by the pump, etc., is effected by applying somewhat excessive cooling, and by compensating the excess with the aid of a variable heater element, which surrenders its heat directly to the suspension. The heater element is controlled by a thermometer, which measures the temperature of the suspension. This method of stabilization is in the present case more rapid and sensitive than one in which the coolant flow is varied. In the state of equilibrium the heater gives up about half its maximum power (5 kW), making it possible to offset variations of more than 2 kW in the power to be dissipated.

The system is represented schematically in *fig. 1*. Cooling takes place in two stages. Between the

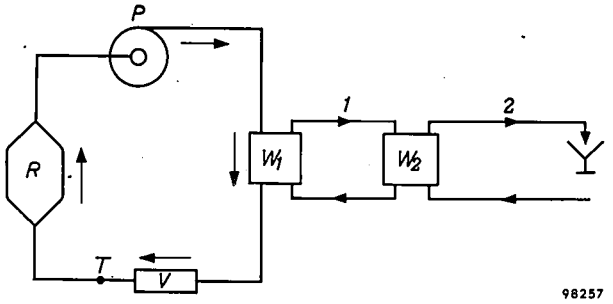


Fig. 1. Schematic representation of reactor circuit with the primary (*1*) and secondary (*2*) cooling circuits. *P* pump. *R* reactor vessel. *V* heater. W_1 primary cooler. W_2 secondary cooler. *T* thermometer in control system.

coolant pipeline *2* (through which main water flows that passes out into the drains, i.e. ultimately into the Rhine) and the reactor circuit, there is a closed-circuit coolant system, which excludes the possibility of radioactive substances being discharged into the river. The closed circuit is called the *primary* cooling system, the other the *secondary* cooling system.

It will also be noted that the thermometer element for the control system is not situated inside the reactor vessel — whose temperature is the decisive quantity — but in the line between the heater and the vessel. The choice of this position was prompted

*) Formerly with the R.C.N.-KEMA team, Arnhem, The Netherlands.

[1]) See article I of this series, J. J. Went, Instrumentation for a subcritical homogeneous suspension reactor, I. Reasons behind the choice of a homogeneous suspension reactor, Philips tech. Rev. 21, 109-121, 1959/60 (No. 4/5). Designed and supplied by Philips.

by the following considerations. Owing to the relatively large volume of the reactor vessel there is a fairly long time-interval between the moment at which the heater delivers too much (or too little) heat, and the moment at which a thermometer situated in the reactor vessel responds to the change. The rate of flow in the vessel is low and one cannot always be sure that the contents are thoroughly mixed. Of course, this time interval can be shortened by stirring, thus limiting it virtually to the time taken by the fluid to pass through the pipe between heater and vessel (the distance-velocity lag). A further small improvement is gained by using a highly sensitive thermometer (or by increasing the loop gain). Even then, however, the interval will still be fairly long. The combination of a long distance-velocity lag and a high loop gain can easily make the circuit unstable. Placing the thermometer immediately after the heater shortens the lag sufficiently to make it easy to keep the control system stable.

The fact that it is not the temperature in the reactor vessel that is kept constant, but the temperature at the output of the heater, is no objection in the operating conditions involved here. In the first place the suspension does not remain long enough in the reactor vessel to cool down appreciably — the fluid is circulated twice to four times per minute — and secondly in a subcritical experiment virtually no energy is generated in the reactor vessel by the nuclear fissions.

In one respect it is even an advantage to place the thermometer before the reactor vessel. Since, owing to mixing, a degree of temperature equalization takes place in the vessel, and the heat capacity of the whole contents is relatively high, rapid periodic variations in the temperature of the influent do not significantly affect the temperature in the vessel. This means that the control circuit (now consisting of a short section of pipeline, a thermometer and an apparatus via which the thermometer controls the heater) need not be critically damped; a possible damped oscillation of the control system has practically no effect on the temperature in the vessel. If the reactor vessel were an integral part of the control circuit, critical damping would be essential.

It may be concluded from the foregoing that the control system is required to do no more than compensate for the variations in the power dissipated in the fluid by the pump and the variations in the power withdrawn by the cooler (which are a consequence of variations in the temperature of the water in the secondary cooling circuit). These variations are always slow and cause a temperature change in the reactor circuit of no more than a few tenths of a degree per hour. Thus, the disturbances to be eliminated are these slow temperature fluctuations of the water entering the heater. The heater power need not be varied more than from 0 to 5 kW for this purpose. Only if the temperature of the cooling water goes on rising or falling during a whole working day may the variability of the heater power be found to be inadequate. In that case it will finally be necessary to alter the setting of the coolant valve. This does not significantly detract from the effectiveness of the control system. There is ample time for such an intervention and moreover it is seldom necessary.

In principle a much lower power would suffice (e.g. 2 kW). A maximum heater power of 5 kW was chosen, however, because the heater is used not only for control purposes but also for heating the suspension prior to the experiment. Moreover, the relatively wide limits within which the heater power is variable make the setting of the cooling capacity less critical.

The temperature changes corresponding to the power contributions of pump, cooler and heater are obviously related to the volume of suspension flowing past any point per unit time. In normal conditions this is 7.5 m^3 per hour. The pump then supplies a power of about 5 kW, the heater an average of 2.5 kW (max. 5 kW), giving 7.5 kW in all. The temperature difference between the inlet and outlet of the cooler is about 1 °C, and between the input and output of the heater 0.7 °C at the most.

Control system

The suspension is heated by simply passing an electric current through part of the stainless-steel pipeline which, together with the reactor vessel, forms the reactor circuit. This method has the virtue of requiring no leads through the pipe wall. The section of pipe serving as heater element measures about 1 metre in length and has an inside diameter of 32 mm. Its electrical resistance is 1.4×10^{-3} Ω, so that for a heat production of 5 kW the current through the pipe wall must be 1900 A. This current is obtained by means of a specially designed transformer (primary voltage 220 V, secondary voltage 3.5 V, secondary current 2500 A max.). The secondary winding is joined to the ends of the heater element via flexible copper strips brazed to copper blocks (*fig. 2*).

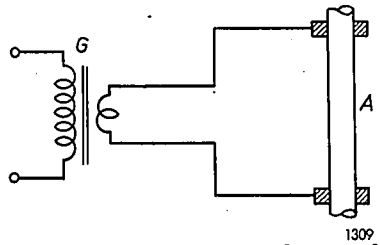


Fig. 2. Heating is provided by passing the secondary current of a transformer (max. 2500 A) through the wall of a section of pipe forming part of the reactor circuit. The transformer *G* is connected to the pipe *A* by flexible copper strip brazed to copper blocks.

The heat produced by this element is regulated by varying the primary current of the transformer between 0 and 32 A. The primary current is passed for this purpose through an antiparallel arrangement of two PL 260 thyratrons, whose moment of ignition varies with the temperature of the fluid (*fig. 3*).

The thyratrons are ignited by a voltage pulse applied to their grids; this is superimposed on a permanent grid bias that keeps the valves cut-off in the absence of a pulse. As the anode voltage is

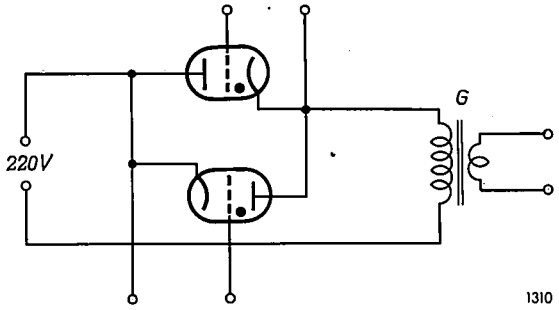


Fig. 3. Primary circuit of heater transformer with antiparallel thyratrons for regulating the current. When the heat output has the correct value, the thyratrons ignite at the moment when their anode voltage is maximum. If the fluid becomes too hot, the moment of ignition occurs later and the current drops; if the fluid becomes too cold, the moment of ignition occurs earlier and the current rises.

sinusoidal and the ignition potential depends linearly on the grid voltage, the "permanent" grid bias can simply be an alternating voltage in antiphase with the mains voltage (*fig. 4*). The method of ignition described makes it possible for the valves to be ignited also in the second half of the time-interval in which the anode voltage is positive, so that the current can also have values lower than half the maximum and even drop practically to zero [3]).

The circuit that supplies the pulse, and which we shall term the ignition circuit, is shown in *fig. 5*. The installation contains two such circuits, one for each thyratron (fig. 3). The heart of each ignition circuit is another thyratron (PL 2 D 21), the anode voltage for which is obtained from the mains via a transformer, and is thus likewise an alternating voltage. The anode current produced upon ignition flows through the primary of the output transformer T_2. As the output voltage of this transformer is proportional to the time derivative of the anode current, T_2 gives a relatively high voltage pulse at the moment this tube ignites.

Of course, the moment at which the thyratrons in the ignition circuits start to conduct must also be

[3]) An example of an earlier application of pulse-controlled thyratrons is given in Philips tech. Rev. **12**, 83-93, 1950/51.

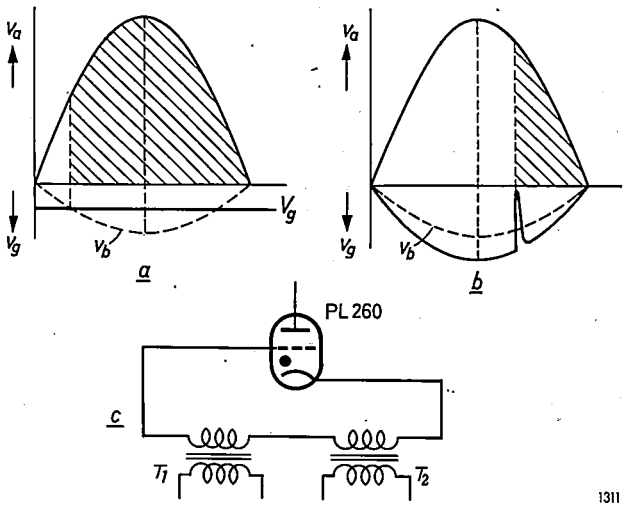


Fig. 4. *a*) If the anode voltage v_a of a thyratron varies sinusoidally, so does the value v_b of the (negative) grid voltage which is just sufficient to prevent ignition. Where v_g has a constant value V_g which is smaller than the maximum value of v_b, the thyratron ignites at the moment that V_g is less negative than v_b and remains conductive for the remainder of the half-cycle (of mains frequency). Ignition in the second half of this time-interval is impossible with a fixed grid voltage. *b*) Ignition in the second half of a half-cycle is possible if v_g consists of a sinusoidal voltage greater than v_b (and in antiphase with v_a) with a pulse superimposed on it.
c) The sinusoidal component of v_g and the pulse are fed to the grid via separate transformers T_1 and T_2.

capable of displacement within a whole half-cycle of the mains voltage. This is achieved by making their grid voltage a composite one consisting of:

a) an alternating voltage v_1 lagging 90° in phase behind the anode voltage v_a;
b) a variable direct voltage V_2;
c) an alternating voltage v_i whose amplitude V_i is proportional to the difference between the measured and the desired temperature of the

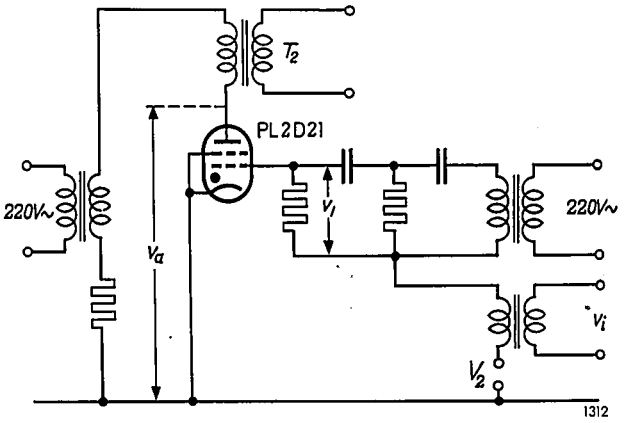


Fig. 5. Diagram of one of the two ignition circuits. v_1 alternating voltage lagging 90° in phase behind v_a. v_i alternating voltage whose amplitude varies with the temperature of the suspension and which is either in phase or in antiphase with v_a, depending on whether this temperature is too high or too low. By varying V_2, a direct voltage, it is possible, when the amplitude of v_i is zero, to make the tube ignite at the precise moment when v_a is maximum.

suspension. This voltage is in phase or in antiphase with v_a, according to whether the suspension is too cold or too hot, and is derived from a resistance thermometer.

Fig. 6 shows how v_g varies with time and how the variation of V_i can shift the moment at which the pulse occurs over almost the whole time-interval of a half-cycle of the mains voltage. The direct voltage V_2 serves for making the moment of ignition when $V_i = 0$, that is when temperature equilibrium has been reached, coincide with the moment at which v_a is maximum. The power

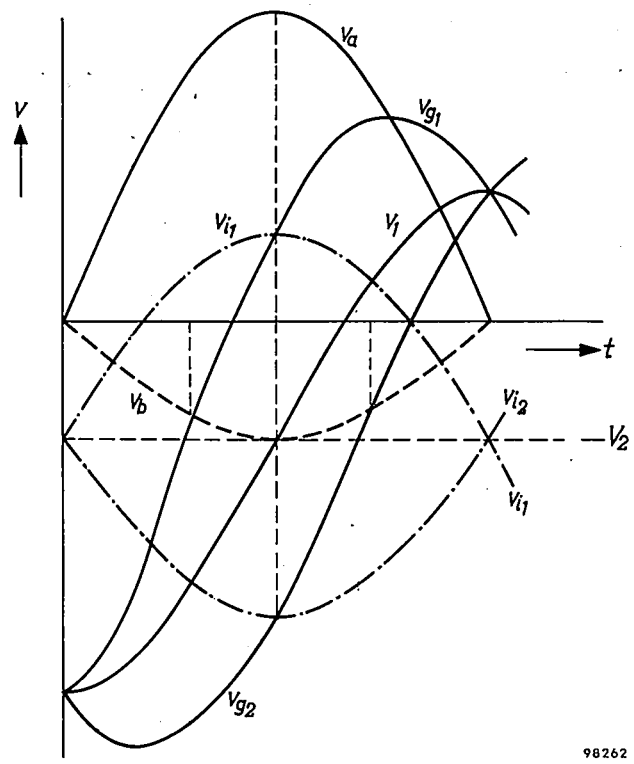


Fig. 6. Components of the thyratron grid voltage in one of the ignition circuits. The voltage v_i is denoted by v_{i1} when the suspension is too cold and by v_{i2} when it is too hot. The curves for v_1, v_{i1} and v_{i2} are drawn about the zero-line $v = V_2$. Curves v_{g1} and v_{g2} represent the time variation of the grid voltage $V_2 + v_1 + v_i$ when v_i has the two waveforms as shown. These waveforms correspond to 0.02 °C too low and 0.02 °C too high, respectively. Since v_i can have a much larger amplitude than represented here, the point where the v_g and the v_b curves intersect can be shifted over almost the entire time-interval of a half-cycle.

dissipated in the heater element, and the average current, then have half their maximum values. It can be shown that the power delivered by the heater to the suspension depends, over a considerable temperature range, more or less linearly on the temperature difference to be corrected (*fig. 7*).

The resistance thermometer mentioned is connected conventionally in a Wheatstone bridge (*fig. 8*). The temperature element itself is a platinum wire R_T, whose resistance at 20 °C is about 100 Ω. The bridge is fed by a 10 V alternating voltage of mains frequency. The comparison resistance R_R is variable in 20 steps each corresponding to 4 °C, and further into 20 steps each corresponding to 0.2 °C. The error voltage appearing between a and b, representing the temperature deviation from the desired value, is about 1 mV per 0.1 °C. This voltage is too small to serve as v_i (figs. 5 and 6) and is therefore considerably amplified before being applied to the ignition circuits. The amplification is such that a deviation of 0.05 °C from the desired temperature causes the moment at which the ignition circuit delivers a pulse to shift almost to the beginning or to the end of the half-cycle concerned, thus switching the heater fully on or off, depending on whether the suspension is too cold or too hot.

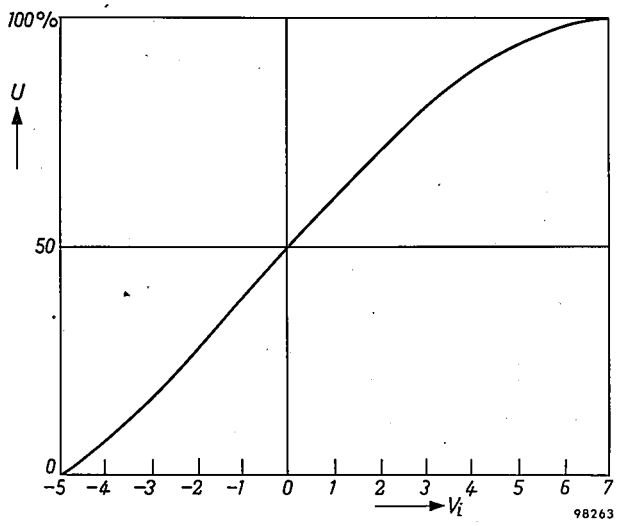


Fig. 7. The power U delivered by the heater varies linearly, over a large part of the control range, with the temperature deviation of the suspension, i.e. with v_i. The value of U is given as a percentage of the maximum; the units along the horizontal axis are arbitrary.

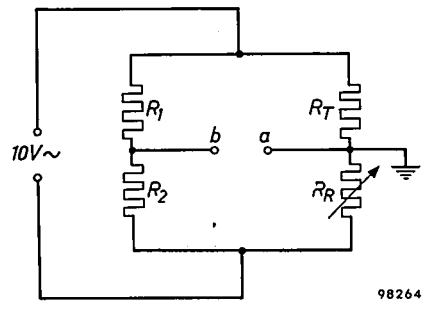


Fig. 8 .Wheatstone bridge circuit for resistance thermometer. R_T platinum wire in thermal contact with the suspension. R_R variable comparison resistor. R_1 and R_2 fixed resistors. The voltage between a and b is amplified and used (as v_i) to control the ignition circuits. R_R can be varied in 20 steps corresponding to 4 °C each, and in 20 steps corresponding to 0.2 °C each.

The amplifier is so designed that, even during the running-up period prior to a reactor experiment, when there is a big difference between the actual and the desired temperature (and hence a relatively high alternating voltage between *a* and *b*), the output signal always remains balanced. Thus, the contributions made by the two PL 260 thyratrons (fig. 3) to the primary transformer current are equal under all conditions. Any asymmetry in this current (i.e. the presence of a D.C. component) could seriously damage the transformer.

A block diagram of the whole control system is given in *fig. 9*.

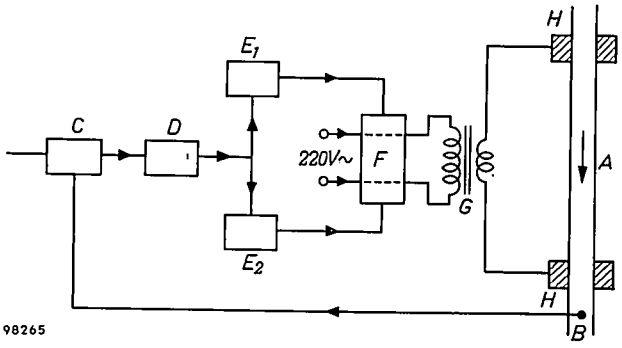


Fig. 9. Block diagram of temperature control system. *A* heater element (section of pipeline). *B* platinum resistance thermometer. *C* Wheatstone bridge, of which *B* is an integral part. *D* amplifier which produces from the bridge voltage a signal capable of driving the ignition circuits E_1 and E_2. *F* regulator of primary transformer current (see fig. 3). *G* heater transformer. *H* connections of transformer secondary to heater element.

Setting the temperature

The first step in setting the temperature of the suspension is to adjust the comparison resistance in the electrical thermometer (R_R in fig. 8) to the value that corresponds to the desired temperature T, and then to switch-on the pump and the control apparatus. The coolant valve remains closed or is only slightly opened. Since the bridge is initially far from balanced, the heater delivers its maximum power. The heat supplied is then about 12 kW (5 kW from the heater and about 7 kW from the pump). As the fluid warms up at the rate of approximately 2 °C per minute, the desired temperature is quite quickly reached. As soon as it is exceeded by 0.05 °C, the heater is automatically switched off (see above). This can be observed on the control panel. The heat production will generally still be excessive, but the rate at which the temperature now rises is evidently greatly reduced (to less than 1 °C/min). Consequently there is ample time to open the coolant valve until the heat extraction reaches, say, 6 kW. The temperature then drops slowly. As soon as it falls below $T + 0.05$ °C, the regulator comes into operation and the temperature is stabilized somewhere in the region from $(T + 0.05)$ to $(T - 0.05)$. The exact value of the equilibrium temperature is determined by the extent to which the coolant valve is opened. The regulator can be set to the middle of the control range by adjusting the opening of the valve while watching the reading on a meter that indicates the power delivered by the heater. This being done, the heater can correct variations of $2\frac{1}{2}$ kW in the coolant or pump power, such that the resulting variations in the suspension temperature are reduced to less than 0.05 °C.

Properties of the control system

We shall now consider the extent to which the system reduces a disturbance and whether the system is stable. To answer both these questions we must know the magnitude of the total gain $A(\omega)$ in the control loop. By this is meant the amplification which a sinusoidal signal of angular frequency ω undergoes after passing successively through all elements of the control loop. In order to determine the function $A(\omega)$ we shall have to cut the loop open somewhere. If all elements are linear, it is immaterial where we do this. Furthermore we must know the way in which the phase shift φ between the input and output signal of the cut-open loop varies with the frequency of the input signal. In short, we need the frequency response (amplitude and phase characteristics) of the open loop.

Fig. 10a and *b* show these characteristics for the system described here. The relation between the gain A and the phase shift φ is represented in a Nyquist diagram as in fig. 10*c*. The length of the line from the origin to a certain point of the diagram is equal to the gain obtained at the frequency in question, and the angle between this line and the axis is equal to φ. The gain is therefore represented in this diagram by a vector.

The vector P by which a control system reduces an incoming periodic disturbance is equal to the absolute value of the vector sum of the vector just referred to, for the frequency of the disturbance signal, and the negative unit vector. Its magnitude, then, can be easily found graphically from the length of the line joining the relevant point of the Nyquist diagram to the point —1 on the axis. At frequencies where $\varphi \approx 0$ the vector sum becomes the arithmetic sum [4]). We can deduce from fig. 10*a* that P is here equal to eight (seven plus one) at such frequencies. *Fig. 11* shows how the reduction factor varies with increasing frequency. For disturbances having an

[4]) The degree to which periodic disturbances are reduced is sometimes expressed as the *deviation ratio* $1/P$.

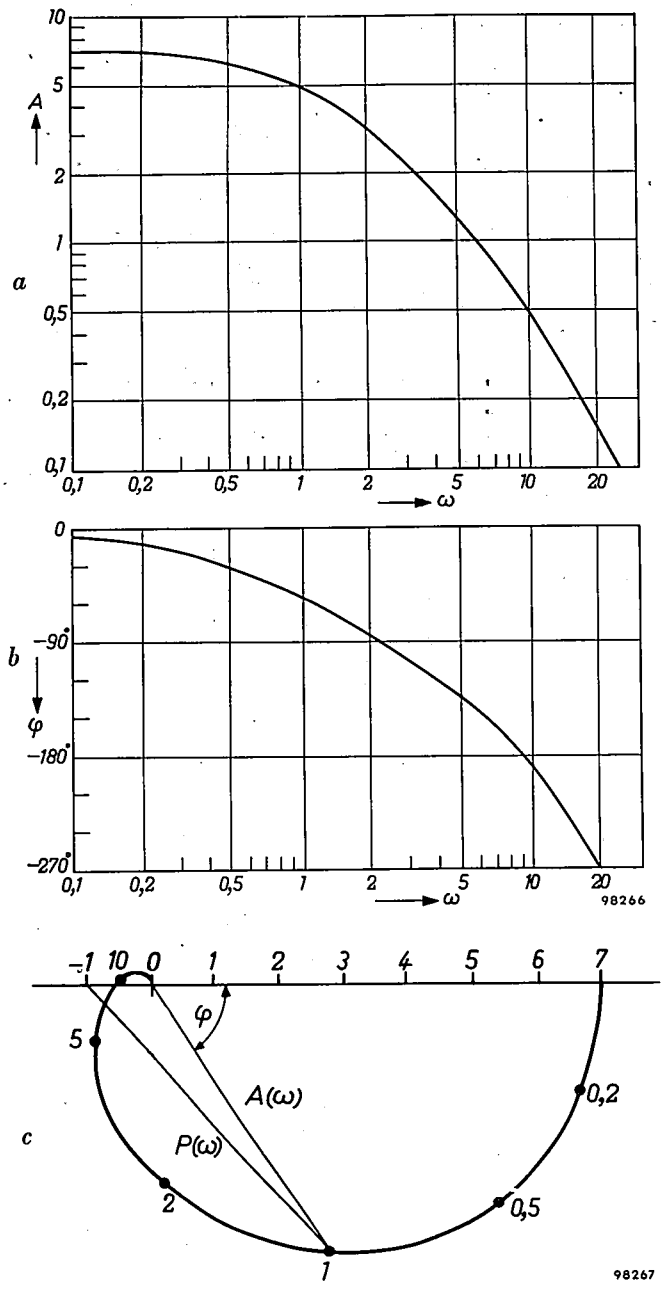


Fig. 10. Frequency-response curves of opened control loop.
a) Amplitude characteristic.
b) Phase characteristic. At very low frequencies the gain is about 7. At the angular frequency ω where the phase shift φ is 180°, the gain is less than unity.
c) Nyquist diagram of temperature control system. The figures around the curve indicate the angular frequency to which the points marked correspond. The distance from these points to the origin is equal to the gain obtained at that frequency; φ is the phase shift. The curve does not enclose the point —1 on the axis (cf. fig. 10*a* and *b*); this implies that the system is stable. The length of the line joining a point on the curve to the point —1 gives the value of the reduction factor $P(\omega)$.

angular frequency between 0 and $^1/_2$ (i.e. $\sim {}^1/_{12}$ c/s) it is seen that P has approximately the value 8. At $\omega = 1.5$ (i.e. $\sim {}^1/_4$ c/s) the reduction factor P has dropped by half. Since, as mentioned, the disturbances occurring in practice are very slow-moving, this "bandwidth" of the control system is entirely adequate.

With regard to stability, fig. 10 shows that the gain is less than unity for the frequency at which φ has increased to 180° — i.e., at which the negative feedback in the loop has changed to positive feedback. From the fact that the Nyquist diagram does not "contain" the point —1, it may be concluded that the control system is indeed stable [5]).

It can also be seen that an appreciable increase in gain, and hence in P, is not possible without endangering the stability. In any case, a higher reduction factor is not needed. A change of no less than 1 kW in the coolant or pump power corresponds to a temperature change, without control, of about $^1/_4$ °C, and with control to a temperature change of only $^1/_{30}$ °C.

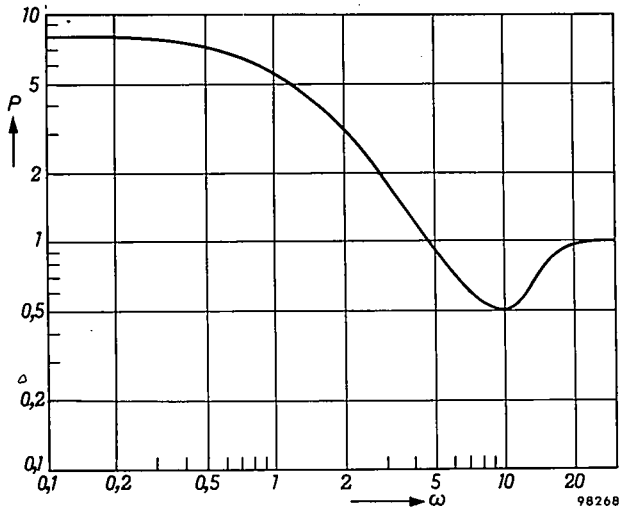


Fig. 11. The factor P, by which periodic disturbances are reduced, as a function of the angular frequency ω of the disturbance (assumed to be sinusoidal).

Measuring the average concentration of the suspension in the reactor vessel

In investigating the properties of a suspension-type reactor, one of the things to ascertain is the relation between the concentration of the suspension and the reactivity, which varies with the concentration in a very marked way. It is therefore

[5]) For complex control systems the stability criterion as formulated here is over-simplified, but for relatively straightforward systems, like the present one, it is adequate. The stability problem is treated in detail in G. S. Brown and D. P. Campbell, Principles of servomechanisms, Wiley & Sons, New York 1948, and in G. J. Thaler and R. G. Brown, Servomechanism analysis, McGraw-Hill, New York 1953.
Editor's note: A derivation of the formulae from which figures 10 and 11 are obtained will be given in a forthcoming article in this journal, dealing with control theory.

necessary to measure this concentration in the reactor vessel as accurately as possible. Unfortunately this cannot be calculated from the quantities of solid particles and liquid contained in the whole reactor circuit. Since a suspension settles, the concentration in a section of pipeline in which the liquid flows upwards will be greater than elsewhere. If the flow rate in such a section is so slow as to equal the rate at which the particles settle (with respect to the liquid), all solid particles will ultimately accumulate in that one section. In the reactor vessel, where the flow rate is slow because of the relatively large diameter, this accumulation effect is clearly perceptible. To determine the concentration in the reactor it is therefore necessary to adopt a method that measures solely the contents of the vessel.

We shall deal here with two methods developed for this purpose. Only one of them is universally applicable. The other, which we shall discuss first, can be used in only one of the two types of reactor vessel appropriate for a suspension reactor. The shape of these two types is dictated by the behaviour of the suspension; in unsuitably shaped vessels the suspension settles on to parts of the walls and there may be considerable differences in concentration from one region to another. Experiments have shown that the suspended particles are most uniformly distributed over the volume and do not settle locally on the walls if the reactor vessel has roughly the shape represented in *fig. 12a*, which, for convenience, we shall call cylindrical. The fluid enters the vessel at the bottom, where rotary movement is imparted to it by the vanes *1*, and leaves at the top, where flow-straightener vanes *2* eliminate the rotary movement. The second type, which we shall call spherical, is shown in fig. 12*b*. This has certain constructional advantages, and the powerful turbulence of the fluid removes the need for special measures to prevent settling of the particles on the walls. The distribution of particles over the volume, however, is not so uniform as in the cylindrical vessel.

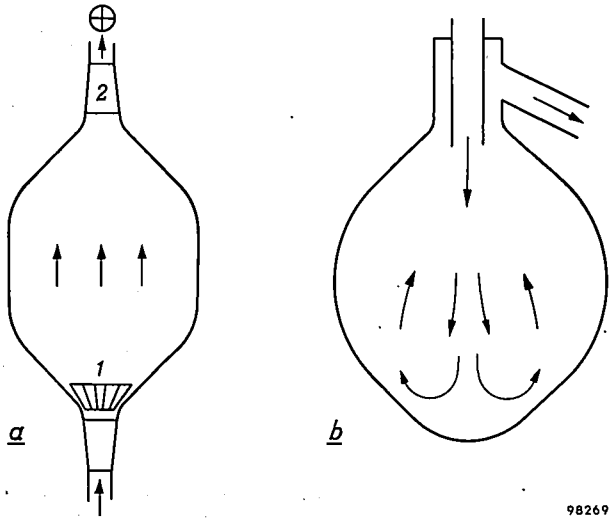


Fig. 12. *a*) "Cylindrical" type of reactor vessel. The vanes *1* impart to the influent a rotary motion, which is eliminated in the effluent by the vanes *2*.
b) "Spherical" type of reactor vessel. The arrows indicate the direction of fluid flow.

An investigation of this uniformity was made by measuring the concentration at various places by means of the absorption of light. The principle of this method is represented schematically in *fig. 13*. Since UO_2 particles absorb virtually all the light, even when the concentration is low and the slit narrow, the measurements were carried out on a suspension of grains of sand of suitable size. *Fig. 14* shows the result of two series of measurements at various heights in a cylindrical vessel.

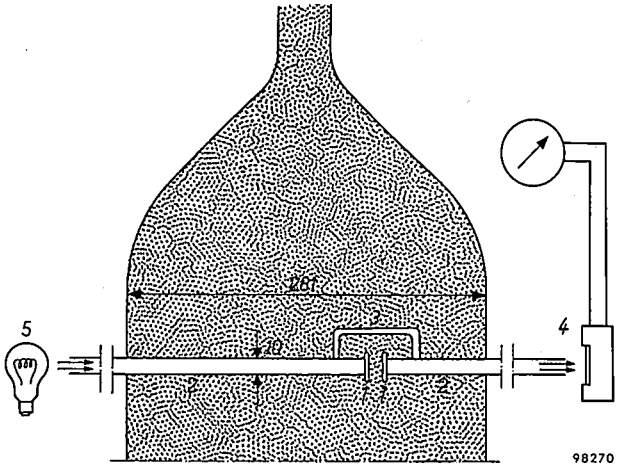


Fig. 13. The local concentration of a (non-radioactive) suspension in the reactor vessel was measured at various heights by means of light-absorption. The windows *1* fitted in the tubes *2* are kept a constant distance apart by brackets *3*. The whole device can be shifted lengthwise. The photocell *4*, connected to a galvanometer, measures the extent to which the light from the incandescent lamp *5* is absorbed.

The direct method

Both methods of measuring the average concentration in the reactor vessel are based on the law that the difference in hydrostatic pressure of two points at unequal heights is proportional to the density of the liquid. The simpler method is illustrated in *fig. 15*. Above and below the reactor vessel, branch arms are led out and connected to a differential manometer. As there are no suspension particles in the manometer tubes — the liquid is at rest and the few particles that do get through will in any case soon settle — the pressure difference is found from the product of the difference in height δh of the liquid levels in both arms and the density ϱ_l of the liquid. Assuming that the flow of the suspension makes no contribution to the

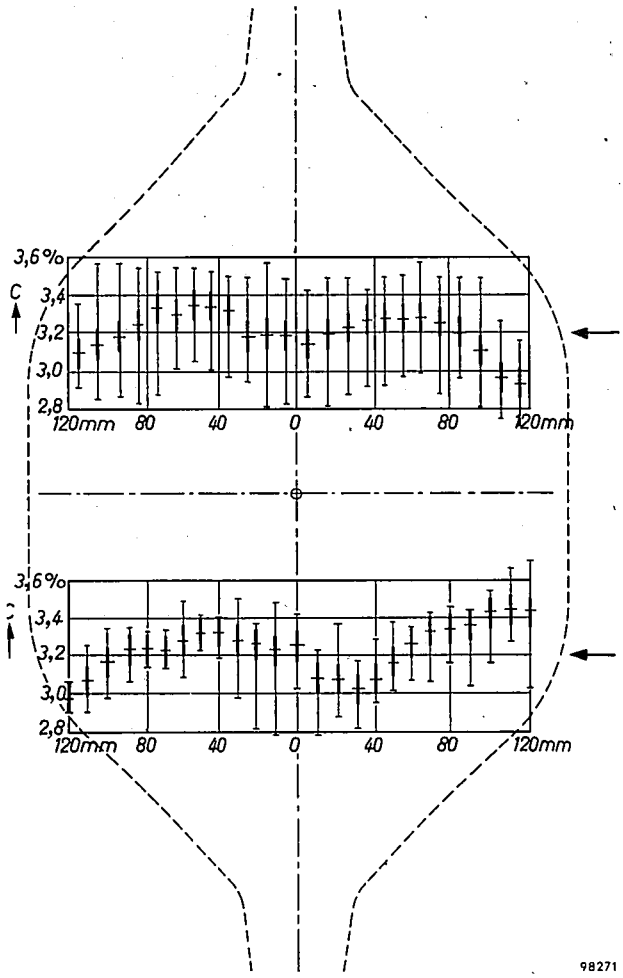


Fig. 14. The concentration, in percentage by volume, of solid particles at various places in the cylindrical reactor vessel (dashed line). The arrows (right) indicate the heights at which the measurements were made. The length of the thick vertical strips gives the standard deviation, the horizontal strips represent the mean values observed. The ends of the thin vertical strips denote the maximum amplitude of the fastest fluctuations followed by the galvanometer (response time 0.5 sec). It is thus possible that fluctuations of even shorter duration and greater amplitude occurred, but these are presumably of no nuclear-physical importance.

pressure difference, then the latter must be equal to $H \times C(\varrho_s - \varrho_l)/100$, where H is the vertical distance between the two manometer connections (fig. 15), C is the concentration of the suspension in % volume and ϱ_s is the density of the solid particles suspended in the liquid. Equating both expressions for the pressure difference makes it possible to find C.

For the maximum value of C, δh is only about 10 cm. Consequently, if a liquid manometer as in fig. 15 is used, its accuracy, which is not very high, will determine the relative error in the result of the measurement. In reality, therefore, the pressure difference is measured with an electrical instrument specially designed for the purpose, viz. a diaphragm-type differential pressure transducer. The sensitivity and very high stability of this instrument (both as regards its zero point and calibration) make it possible to measure pressure difference with a high degree of accuracy [6]).

In general, the flow of the suspension will of course make some contribution to the measured pressure difference. In turbulent flow this contribution is approximately proportional to ϱv^2, where ϱ is the density of the suspension and v is the flow velocity of the liquid. The validity of this relation suggested the means of largely compensating for the disturbance due to flow. If a liquid flows through a pipe in which there is a constriction, the pressure on the pipe wall at the constriction is

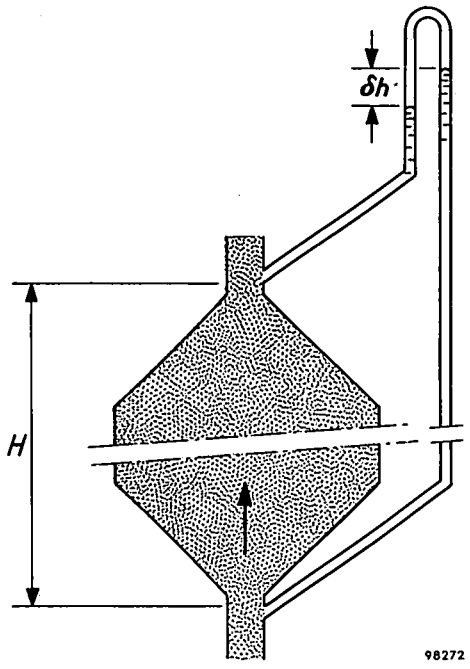


Fig. 15. Principle of the method of measuring the average concentration of the suspension in the vessel from the hydrostatic pressure difference. The liquid in the manometer arms contains no UO_2 particles.

lower than elsewhere, and by an amount that is likewise proportional to ϱv^2. Thus, by applying the upper connection of the manometer to an appropriate point on the reactor vessel — i.e. to a point where the diameter is larger — instead of to the outlet pipe, it was possible to offset the one effect by the other and so to reduce substantially the correction necessary on the manometer reading. The magnitude of the remaining correction was determined by passing the pure liquid through the reactor circuit and measuring the pressure differences at

[6]) This instrument was developed under the direction of J. A. H. Kersten (R.C.N.-KEMA team). It operates on the same principle as commercially available differential pressure pick-ups.

various rates of flow. *Fig. 16* shows a scale drawing of the cylindrical reactor vessel employed in the subcritical experiments, indicating the position and form of the manometer connections.

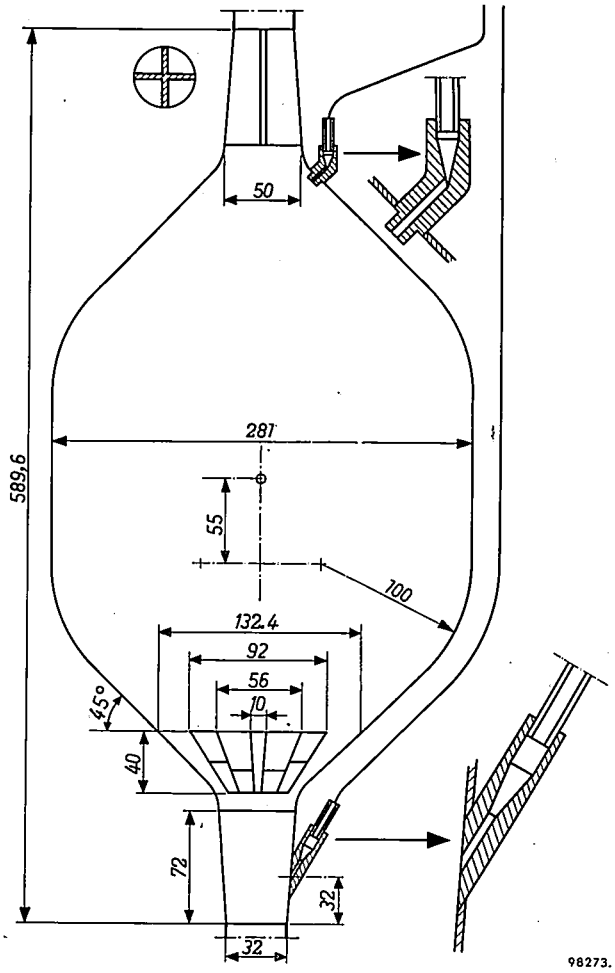


Fig. 16. Scale drawing of cylindrical reactor vessel used in the subcritical experiments, showing the pressure tappings for the differential manometer with which the concentration of the suspension is measured.

The following advantages are associated with the method just described: 1) the equipment required is simple; 2) scarcely any material is introduced on or in the reactor vessel (that would mean extra neutron losses); 3) components that might need repair in the foreseeable future are not positioned close to the reactor vessel (where they would become highly radioactive and therefore impossible or very difficult to repair). A drawback is that the manometer shows some jitter owing to the whirling motion of the liquid in the vessel. That is why this method can only be used for measurements on the cylindrical reactor vessel and not on the spherical type. The violent turbulence in the latter makes the manometer so unsteady that even a moderately accurate reading is quite impossible. In measurements in the cylindrical vessel the uncertainty is about 0.04 % by volume UO_2. Since, under normal conditions, the UO_2 concentration is roughly 2 % by volume, this value corresponds to a relative error of 2%.

The "flow simulator" method

The second method, which is a variant of the first, is distinguished mainly by the comparative steadiness of the manometer and by the fact that the liquid-flow contribution to the pressure difference is practically reduced to zero. This method is accordingly suitable for measurements on a spherical reactor vessel.

The favourable features mentioned are due to the manometer not being connected to the reactor vessel itself or to the reactor circuit, but instead to a device called a "flow simulator", which is in fact a flow shunt connected across the reactor vessel. The form of the simulator can be seen in *fig. 17a*, and fig. 17*b* shows how the device is incorporated

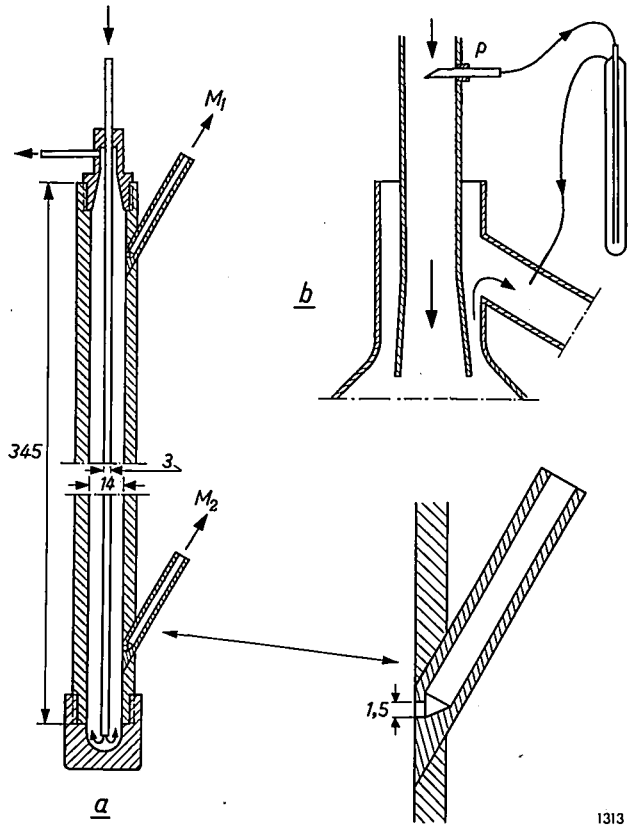


Fig. 17. *a*) Flow simulator. The dimensioning is such that the fluid rises in the wide outer tube at approximately the same velocity as in the cylindrical reactor vessel. The differential manometer is connected at M_1, M_2.
b) Simulator shunted across reactor vessel. For the orientation shown of the pipe p, where the supply line of the simulator joins the reactor circuit, the meter indication is proportional to the average concentration in the vessel and also (within a certain range) independent of the flow rate. At other orientations of p this is not so, owing to the momentum of the larger UO_2 particles at the mouth of the pipe.

in the reactor system. The simulator line is connected to the reactor circuit in front of and behind the reactor vessel. The fluid enters the simulator through the narrow inner tube and then rises in the outer tube, to which the manometer leads are connected. The flow velocity in the outer tube is very low, so that the pressure difference it causes is extremely small. Moreover the flow is very steady, and so does not unsettle the manometer. The length and cross-section of the simulator tube are chosen so that the fluid rises at the same rate as in the reactor vessel. The effect of the flow rate on the concentration in the reactor vessel and simulator is the same and there is thus a proportional relation between the concentration measured on the simulator and that of the suspension in the vessel.

An incidental advantage of the simulator method of measurement is that the vertical distance between the manometer connections — i.e. the sensitivity of the concentration meter — can if necessary be made larger than in the first method.

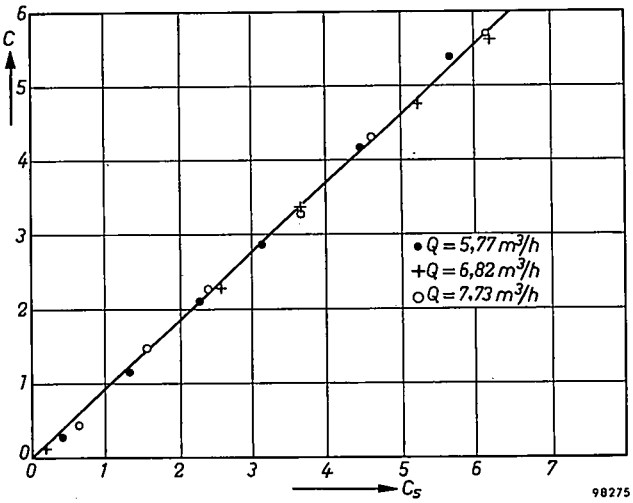


Fig. 18. Calibration of the simulator concentration meter, connected to the spherical reactor vessel. The true concentration (C), determined by weighing the contents of the vessel, is plotted vertically, the concentration (C_s) derived from the meter readings is plotted horizontally. The units on the coordinates are percentages by volume. The meter indication is independent of the flow rate Q in the range of flow rates investigated.

This expedient can only be employed, however, if the concentration does not vary rapidly with time, for it means that the time taken by the suspension to flow through the simulator pipeline will be longer than the time which the fluid through the reactor takes to flow from the first to the second connection point (about 10 sec). Consequently the simulator will not faithfully follow the variation with time of a rapid change of concentration in the reactor vessel.

Fig. 18 shows a result of an investigation into the behaviour of a concentration meter employing a simulator. It can be seen that the indication is admirably proportional to the concentration, and independent of the flow rate of the suspension (at least within the range of flow rates investigated). It also appears, however, that the indication is too high by a constant factor. For absolute measurements this factor must be determined by calibration, but relative changes of concentration can be measured with a good degree of accuracy without calibration.

In the reactor experiments now in progress the reactor vessel is of the cylindrical type. The concentration can be measured by either of the two methods described, although here too the simulator method has proved more accurate. In future investigations of a spherical-type vessel this method, as mentioned, will be used exclusively.

During the initial stage of the investigations a third method was used [7]). On one side of the pipeline close to the inlet of the reactor vessel a radioactive source emitting γ rays was mounted, and on the other side of the pipe a radiation detector. The concentration of the passing suspension was determined from the extent to which γ rays were absorbed. A virtue of this method is that it readily allows an absolute calibration to be made. Drawbacks are 1) that the measurement is not made on the vessel itself; 2) that owing to the limited intensity of the source the measurement takes a certain time before the statistical error in the result is sufficiently small, which means that short-lived changes in concentration cannot be measured, or only very inaccurately; and 3) that the method is difficult to use if the suspension itself shows any marked radioactivity. The latter drawback has already disqualified this method for the reactor experiments now in progress. In the early stages, however, it rendered excellent service as a means of checking the two manometric methods described above.

Measuring the flow rate in the reactor circuit

Among the many methods of measuring the quantity of fluid flowing per unit time through a pipe, the two most widely used are probably the *orifice plate* and the *venturi tube* methods. The principle of both methods is explained in *fig. 19*. In

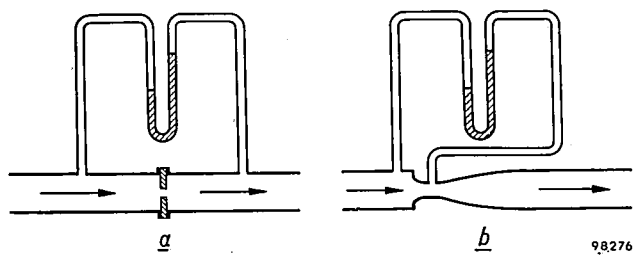


Fig. 19. The two most commonly used methods of flow-rate measurement.
a) The flow rate is determined from the pressure difference over an orifice plate acting as a flow resistance.
b) The flow rate is found from the pressure drop on the tube wall at a point of constriction (venturi tube).

[7]) This method was developed by J. Kalshoven (R.C.N.-KEMA team).

view of the requirements and conditions of the present case, the second method was chosen. A drawing of the somewhat unconventional venturi tube designed for the purpose is shown in *fig. 20*. It will be seen that the diameter at the constriction is not much smaller than elsewhere, so that the tube introduces no appreciable flow resistance into the circuit. Moreover, wear due to abrasion by the suspended particles (in the concentration used) is not severe, and therefore the calibration of the meter remains reliable for a long time. The tube is short, as necessitated by the available space. The holes for the leads to the pressure gauge are arranged so as to prevent them from getting clogged by suspension particles.

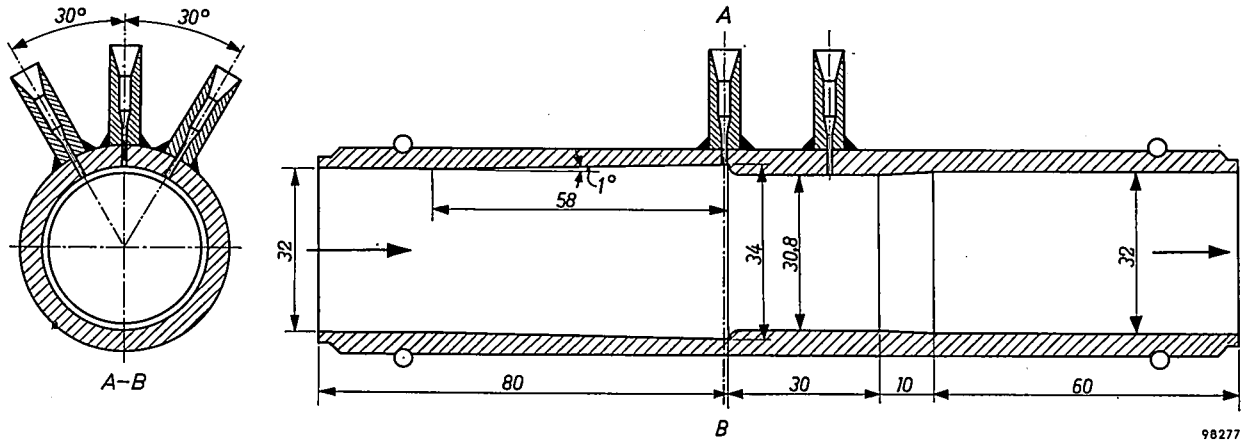


Fig. 20. Scale drawing of venturi tube designed for the reactor. The tube is short and also has a low flow resistance. The outlets of the manometer pipes, which are usually arranged around the tube to obtain an averaging effect, must not in this case depart more than 30° from the vertical, in order to avoid penetration and possible clogging by UO_2 particles.

In designing the tube the aim was to ensure that the pressure difference would be about 50 cm of water at full pumping capacity, i.e. at a delivery of 10.5 m³ per hour. This is in fact the case. Calibration is carried out using clean water and an orifice plate for comparison. The calibration applicable to a suspension is found from that applicable to water by multiplying the meter reading corresponding to a given delivery by the density of the suspension [8]). The pressure difference is measured with a Philips differential pressure transmitter, type PR 6000 F; it is recorded on a type PR 2200 recorder.

Survey of the whole reactor system and instrumentation

We shall now review the whole reactor system together with the other instruments for measuring operating parameters which we have not yet discussed. A layout of the complete installation is given in *fig. 21*. The pipelines and vessels containing the suspension are shown red. Those outlined in black represent the reactor circuit, the others are ancillary lines connecting this circuit to fuel (UO_2) removal and replenishment plant. The primary coolant circuit, containing distilled water, is shown blue, and the secondary, through which mains water flows, is green. The other blue pipeline also contains distilled water which is in addition free from CO_2; this is fed to the pump to "flush" the bearings. Since this water is slightly under pressure, it forces its way slowly in and prevents the penetration or leakage of radioactive suspension particles. The yellow lines contain compressed air (2 atm) which is used, among other things, for emptying the dump tank (see figure) and for the bubble pump, which keeps the water moving in the primary coolant circuit [9]). Not shown is a compressed-air line (4 atm) for operating the valves by remote control.

The flow rate in the reactor vessel can be varied by means of a three-way valve *A* in the reactor circuit. Adjustment of this valve changes the volume flow in the reactor vessel relative to that in the vertical shunt pipe *AB* (shown under the valve). The small volume of water forced into the circuit via the pump-shaft gland can escape via the riser pipe. The flow rate in this pipe is so slow that the UO_2 particles remain at the bottom. Further particulars are given in the caption to the figure.

The principal operating parameters are recorded. These are:

[8]) The pressure difference is proportional to ϱv^2 (see page 128). Strictly speaking the formula is not fully applicable here, since the viscosity of the suspension is not exactly identical with that of pure water. This, however, can be disregarded in the present case (low concentration), a high degree of accuracy not being required.

[9]) The operation of such a pump was recently discussed in this journal in another connection (Philips tech. Rev. **20**, p. 184, 1958/59, No. 7).

1) the temperature of the suspension in the reactor vessel;
2) the flow rate of the fluid in the reactor circuit;
3) the average concentration of the suspension in the reactor vessel;
4) the temperature of the water in the primary coolant circuit on leaving the primary cooler.

The recordings are made with type PR 2200 recorders, installed in the control room. *Fig. 22* shows a typical recorded curve of the concentration (see 3) above).

depicted. The meters mentioned above, together with control valves, are mounted on this panel at their appropriate positions in the diagram (*fig. 23*). Some of the meters are duplicated, the parallel ones being in the room where the reactor is installed. The legend to fig. 21 indicates the meters concerned. The meters for the less important quantities, and those for quantities which need to be measured only very occasionally, are not on the control panel but in the reactor room. These include meters for the p_H and the electrical conductivity of the water in

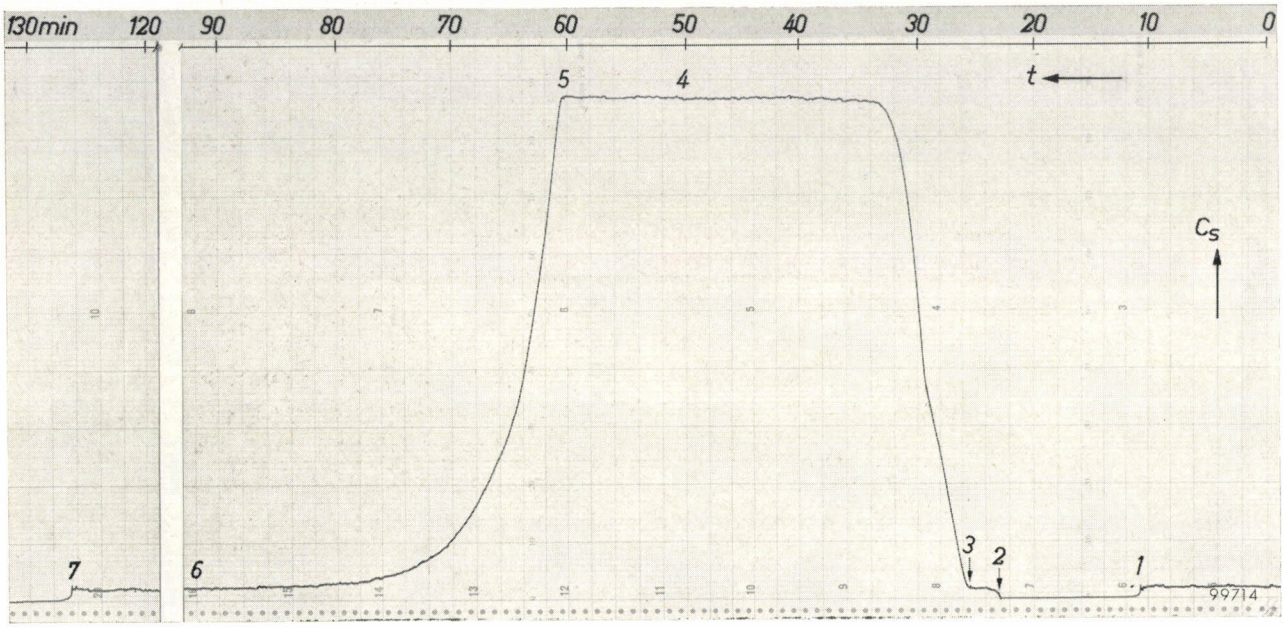


Fig. 22. Recording of the average concentration C_s in the reactor vessel, as a function of time, during various operations. The time scale runs from right to left. Extreme right: pure water is circulating. At *1* the pump was stopped and at *2* re-started: the zero line is unaffected. At *3* UO_2 from the storage vessel was transferred to the reactor vessel; at *4* the concentration has reached an equilibrium value ($\sim 2\%$ by volume). At *5* the cyclone came into operation: at *6* the concentration has fallen to zero. At *7* the pump was once more stopped: the zero line is the same as before. The recording demonstrates the high sensitivity and stability of the differential pressure pick-up used.

The values of most other quantities of importance can be read from meters in the control room. These are:

1) the pressure in the reactor circuit before and after the pump;
2) the temperature of the water entering and leaving the secondary cooler;
3) the flow rate of the secondary coolant;
4) the rate of air flow to the bubble pump;
5) the pressures in the compressed-air lines (4 and 2 atm).

The control room contains a large graphic panel on which the reactor flow sheet as in fig. 21 is

which the UO_2 particles are suspended, a thermometer in the secondary coolant circuit near the input of the primary cooler, level gauges, manometers, etc.

The instruments described in this article have made it possible to make a comprehensive study of the properties of the subcritical reactor system and to monitor the events taking place inside the reactor. Every quantity of importance, or of potential importance, can be measured. Monitoring is not confined to the reactor circuit but extends to the cooling system, so that for example any deterioration of the heat exchange in the coolers as a result

of scale formation can quickly be ascertained. From the experiments carried out so far, much valuable data has been accumulated. Apart from the measurements of the first four quantities mentioned in this section, the measurements on the many secondary quantities have proved of great value.

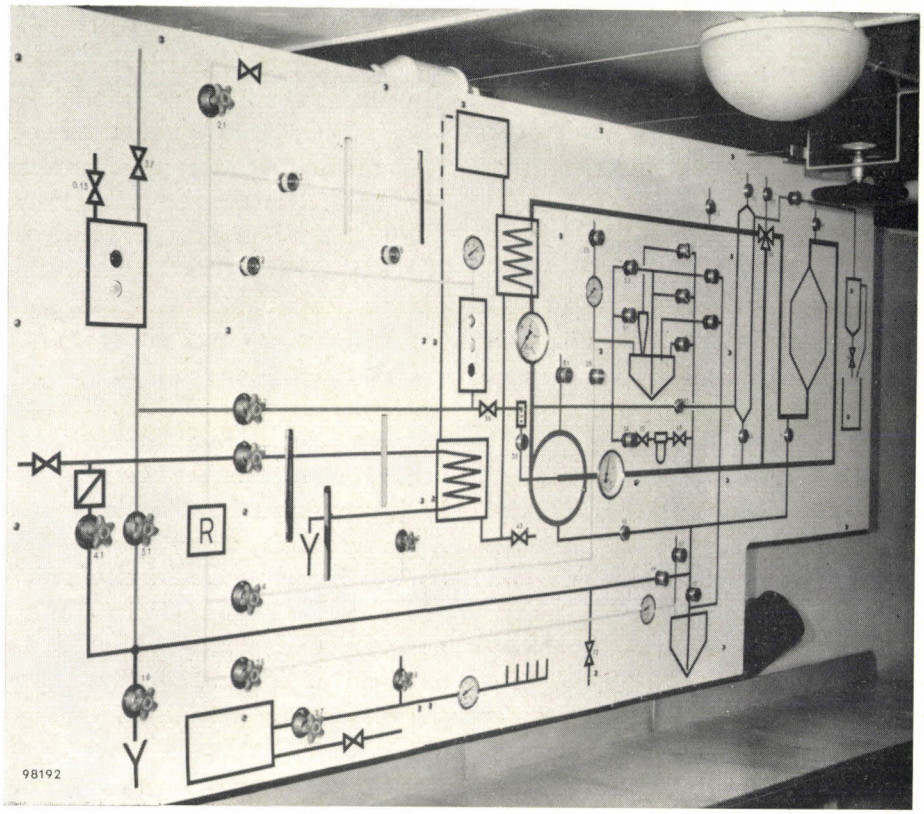


Fig. 23. Graphic panel of the reactor in the control room (cf. fig. 21). The large handwheel (top right) operates the valve that regulates the flow rate through the reactor vessel. Pilot lamps indicate the liquid levels in the storage tank for distilled water and in the pressure-seal water tank. Bottom right is a periscope for reading various meters in the reactor room.

Summary. In experiments on the subcritical suspension reactor at Arnhem it is necessary that the temperature in the reactor vessel should vary by no more than 0.1 °C at any desired value between room temperature and 100 °C. Since the pump that circulates the suspension around the reactor circuit supplies energy to the fluid (about 5 kW), cooling is required at all operating temperatures. As neither the pump energy supplied nor the cooling are perfectly constant, the temperature must be regulated. This is done by means of a heater (a section of pipe in the reactor circuit through which an electric current of up to 2500 A is passed). The heater is controlled by a resistance thermometer and control circuit. The "coolant power" is so adjusted that at the desired temperature the heater delivers half its maximum output (which is 5 kW). The power supplied to the heater is varied by regulating the primary current of the transformer to whose secondary the heater is connected; the regulation is effected by two antiparallel thyratrons through which the primary current passes. These are ignited by a pulse which can be advanced or delayed in both half-cycles of the mains voltage, depending on whether the suspension is too cold or too hot. With this method, the moment of ignition can fall in the second half of a half-cycle of the mains, and thus the current is variable from zero to maximum. Slow variations in temperature are reduced by a factor of 8. Rapid variations cannot occur in practice. The Nyquist diagram of the control system shows that the installation is stable.

The average concentration of the suspension in the reactor vessel is determined from the difference in hydrostatic pressure between two points at different heights. With a vessel of cylindrical shape, in which the fluid flow is relatively steady, the differential pressure gauge can be connected to the reactor circuit itself. A correction must be applied for the pressure difference caused by fluid flow. With a spherical vessel the flow is too turbulent for this method. Here the concentration is measured by connecting the manometer to the wide outer tube of a "flow simulator" shunted across the reactor vessel. The flow rate in this vertical tube is roughly the same as in the reactor vessel, and the concentration in both is affected in the same way by the settling of suspension particles. An additional merit of this method is that the correction for fluid flow is negligible. The flow rate in the reactor circuit is measured with a specially designed venturi tube, having a low flow resistance.

The principal operating parameters are recorded in the control room; many others are indicated on meters. Duplicates of some of these meters are contained in the room where the reactor is installed. Meters for less important or only occasionally measured quantities are not in the control room but only *in situ* on the reactor.

IIIA. THE MONITORING OF LOW NEUTRON FLUX BY MEANS OF FAST PULSE-COUNTING CHANNELS

by J. J. van ZOLINGEN *). 621.039.564.2:621.039.524.46:621.374.32

Introduction

The specific nature of a chain reaction requires that the neutron-flux level in a nuclear reactor should at all times be known. Particularly during the start-up, and even after a shut-down, this quantity and its increase or decrease with time must be under continuous observation.

In a reactor moderated by heavy water the (thermal) neutron flux during start-up is only about 10^{-6} of its rated value, and about 10^{-9} in a graphite-moderated reactor. The neutron-flux measurement must thus cover a range of at least six or nine decades, as the case may be.

Of the other demands to be met by neutron-flux meters the first is a high standard of reliability. There must also be certainty that the indication is not partly caused, or at least not to any significant extent, by the gamma rays always present in a nuclear reactor. Moreover, the neutron-flux detection system must have a rapid response, for in a nuclear reactor the power level can rise extremely rapidly and the automatic safety system or the operating personnel must be able to intervene in good time. Ideally the detection time should be determined only by the statistics of the elementary processes and by the desired degree of accuracy.

There are two ways of measuring neutron flux:

A) by counting the number of pulses which the individual neutrons produce per second in a counter tube, and

B) by measuring the ionization current caused by the neutrons in an ionization chamber.

If both types of neutron detector — counter tube and ionization chamber — have the same volume and gas-filling, they are equally sensitive. Method B has the advantage of simpler equipment, but its use depends on the neutron flux being high enough to cause an accurately measurable ionization current. With method A, on the other hand, the highest neutron flux that can be measured is determined by the rate at which the counting channel is able to count the separate pulses.

The choice between the two methods depends primarily on the intensity of the background (gamma radiation). In the KEMA subcritical reactor the use of a lead shield to attenuate the gamma rays was ruled out owing to lack of space. Method A allows a certain discrimination in respect of pulse amplitude, small pulses caused by gamma rays being electronically blocked and so not counted. With method B the interference from gamma radiation can, if necessary, be compensated by means of a double ionization chamber, both concentric parts of which are sensitive to gamma rays but only one part to neutrons; the difference current of the two parts is then a direct measure of the neutron flux.

Sometimes, however, the latter expedient in method B is not sufficient. The gamma radiation may be relatively so intense that after compensation, imperfect at the best of times, a considerable background remains. This is the case, for example, if the reactor is started up again after a brief shut-down, when there are few neutrons but a high level of residual gamma radiation. In such a case it is necessary to resort to method A, i.e. to count the numbers of discriminated pulses per second, i.e. the *count rate*.

The best solution is often to divide the total measuring range (of six or nine decades) into two ranges that slightly overlap, and to measure low neutron flux by counting pulses, and high neutron flux by measuring the ionization current. This is done in the case of the KEMA subcritical suspension reactor [1]). In this article, IIIA, we shall discuss the pulse-counting method. The ionization-chamber method will be dealt with in article IIIB.

The experiments made with the KEMA subcritical reactor included an investigation of the *fluctuations* of the neutron flux. Since it was not known beforehand what the correlation times of these fluctuations would be, it was necessary to try and observe them in the shortest possible time; an observation is always the average over a certain integration period, and if the latter is long compared with the correlation time, the fluctuations will be largely averaged out again. If, notwithstanding the inevitable statistical fluctuations in the count rate, we wish to detect in a short integration period the fluctuations of the neutron flux, the counting apparatus must be capable of dealing with very high count rates.

*) N.V. KEMA, Arnhem, The Netherlands.

1) J. J. Went, Instrumentation for a subcritical homogeneous suspension reactor, I. Reasons behind the choice of a homogeneous suspension reactor, Philips tech. Rev. **21**, 109-121, 1959/60 (No. 4/5).

Suppose we want to detect neutron-flux variations of 1% in 1 second with an accuracy of ± 10%; the apparatus must then be able to handle a count rate of $n = 10^6$ pulses per second. The relative standard deviation from the number n is $\sqrt{n}/n = 10^{-3} = 0.1\%$. There is then a 50% chance that the observation will differ by less than 0.1% from the true average value, and that an observed deviation of 0.1% is a real deviation.

The KEMA reactor is surrounded by a highly effective neutron reflector, consisting of blocks of beryllium oxide and graphite. In order not to detract unduly from the favourable properties of this reflector, only narrow channels are made in it for the neutron detectors (see fig. 5 of article [1])). There is therefore no room for shielding against gamma radiation. Consequently it is only when the neutron flux is fairly high that method B can be used. The measurement by the counting method (A) should extend at least to the value of neutron flux covered by B, which implies a count rate of the order of millions of pulses per second.

Originally the reactor was equipped with a counting system covering five decades and capable of a maximum count rate of 10^5 pulses per second. Later, in the course of 1955, new types of discriminators were designed in the KEMA laboratory. These, in conjunction with a reliable flip-flop circuit of very fast response, resulted in greatly simplified pulse-counting systems with count rates tens of times higher than the old ones and more than a hundred times more sensitive.

A pulse-counting neutron-flux meter

To measure a neutron flux by the pulse-counting method, use is made of a counter tube containing boron, or of a fission counter. (A third type of neutron detector is the scintillation counter; this can deliver pulses at a very high rate, but being delicate and unable to withstand high temperatures it is not suitable for use in a reactor.)

Of the two kinds of counters [2]) the one most used, because of its high sensitivity, is that containing the boron isotope ^{10}B, often as boron trifluoride gas, BF_3. A thermal neutron causes the following nuclear reaction in ^{10}B:

$$^{10}_{5}B + {}^{1}_{0}n \rightarrow {}^{7}_{3}Li + {}^{4}_{2}He + 2.5\ \text{MeV}. \quad . \ . \quad (1)$$

The liberated energy of 2.5 MeV per captured neutron appears as the kinetic energy of the lithium ion and the alpha particle. Both particles consequently leave behind in the gas a trail of electrons and ions, and these charge carriers give rise to a voltage pulse between the tube electrodes.

A fission counter contains a fissile material, e.g. ^{235}U in the form of a solid compound, in which fissions are brought about by thermal neutrons. (A process giving rise to a light particle, a deuteron or — as in (1) — an alpha particle, is not normally called a fission.) The fission fragments are multiply charged heavy ions of high kinetic energy; these ions and the liberated electrons ionize the gas with which the chamber is filled, resulting here too in a voltage pulse at the output terminals. The fission chamber has a very high count rate and its pulses are easily separated from the background, but it is less sensitive than the BF_3 chamber.

The voltage pulses delivered by the BF_3 chamber range in amplitude from about 0.2 to 10 mV. Before these pulses can be separated with sufficient accuracy from the lower-amplitude pulses due to gamma radiation and other background, they must be amplified. The separation takes place in a *discriminator*, a circuit that produces an output pulse only if the amplitude of the input pulse exceeds a certain minimum value (the discrimination threshold). The discriminator output may be applied to a *pulse shaper*, which produces uniform pulses suitable for the actual counting system to handle.

In practice it is desirable to have two counting systems, one logarithmic and the other linear [3]). The logarithmic instrument gives a deflection proportional to the logarithm of the neutron flux Φ. It must have a scale covering many decades; this is of particular importance when the neutron flux changes steeply, as it does during a start-up. The linear instrument (deflection proportional to Φ) provides a more accurate reading.

Another important quantity in a nuclear reactor is the *period* T, which is the reciprocal of the relative flux change per unit time, $(d\Phi/dt)/\Phi$. The voltage output of the $\log \Phi$ meter is differentiated with respect to time and a voltage is then obtained that is inversely proportional to T; this voltage is applied to a meter on which T can be read off directly in seconds [4]).

A block diagram of the apparatus discussed is given in *fig. 1*.

If the count rate is greater than 10^5 pulses per second, various difficulties may arise, mainly in regard to pulse discrimination and shaping.

---

[2]) W. Abson, P. G. Salmon and S. Pyrah, The design, performance and use of fission counters, Proc. Instn. Electr. Engrs. **B 105**, 349-356, 1958 (No. 22).
W. Abson, P. G. Salmon and S. Pyrah, Boron trifluoride proportional counters, Proc. Instn. Electr. Engrs. **B 105**, 357-364, 1958 (No. 22).

[3]) The reasons for this are outlined in Philips tech. Rev. **19**, 249, 1957/58.

[4]) See e.g. Philips tech. Rev. **19**, 273, 1957/58.

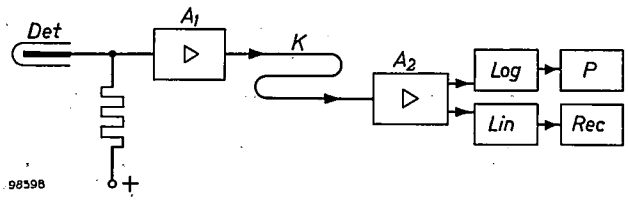


Fig. 1. Block diagram of a two-channel neutron-flux meter and period meter. *Det* neutron detector (e.g. $^{10}BF_3$ counter tube). A_1 preamplifier. *K* cable. A_2 amplifier. *Lin* linear count-rate meter with recorder *Rec*. *Log* logarithmic count-rate meter. *P* period meter.

The "millivolt discriminator"

Drawbacks of conventional discriminators

The pulse-amplitude discriminator hitherto most commonly used is the Schmitt trigger circuit. This is a monostable arrangement of two valves, *I* and *II*; in the quiescent state valve *I* passes current and valve *II* is cut off. A pulse of sufficient amplitude makes *II* conduct and cuts off *I*, after which the circuit returns to its initial state. The minimum pulse amplitude required to trigger the circuit can be preset with a variable biasing voltage.

Drawbacks of this circuit are that the minimum amplitude referred to is fairly large (several volts) and moreover shows a variation of about 5% under low bias conditions. The relative accuracy of the discrimination is therefore not high. It is improved if the biasing voltage is increased to e.g. 100 V, but this calls for correspondingly greater amplification of the pulses. The complication which this involves is considerable, in view of the high gain already required (up to 10^5 times for pulses from a fission chamber) and the wide bandwidth necessary to cope with a high count rate. The result is that it is seldom possible to meet the requirements with less than nine or ten amplifying valves. The fact that the stable state must be rapidly restored after every pulse delivered — otherwise the discrimination threshold would drift — represents an added complication.

The latter is bound up with the "dead time", i.e. the time during which the system after responding to an incoming pulse, is unable to respond to a succeeding pulse [5]). If the dead time is constant, a correction can be applied to allow for the pulses not counted. This is not possible, however, if the dead time varies with the count rate.

These difficulties are largely overcome if a discriminator can be designed whose dead time and discrimination threshold remain constant up to very high count rates, and which is reliably triggered by narrow pulses of the order of 10 mV.

A new discriminator

The fairly low sensitivity of conventional trigger circuits may be explained as follows. In the stable state the cut-off valve has zero transconductance (slope) and hence no gain. The curvature of the I_a/V_g characteristic (i.e. change of slope) is so small that a grid voltage excursion of several volts is needed before the transconductance reaches the value at which the circuit becomes unstable (for this to happen the loop gain is required to be unity). Over this grid-voltage excursion the average transconductance is low and so therefore is the gain; in other words, the valve is not used at the most favourable region of its characteristic.

The situation is improved considerably if a trigger circuit is built in which both valves conduct in the stable state [6]). The principle is illustrated in *fig. 2* (this circuit differs somewhat from that actually used). The input valve *I* is coupled to the output valve *II* via a diode D_1. In the stable state the diode is blocked by the potential drop across resistor R_a (notwithstanding the bias voltage E_1) such that the loop gain (feedback via common cathode load) is less than unity, although both valves are biased to a normal operating point where the transconductance is high. A weak negative pulse on the grid of valve *I* is sufficient, for reasons which we shall

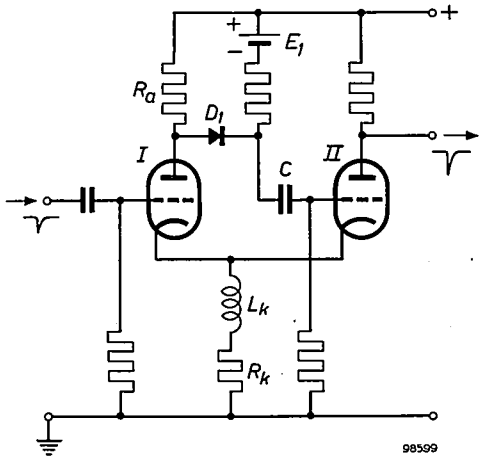


Fig. 2. Principle of "millivolt discriminator". In the quiescent state, valves *I* and *II* are both biased to an operating point where their slope is high, and the diode D_1 is blocked by the voltage drop across the anode resistance R_a, in spite of the biasing voltage E_1. A negative input pulse of sufficient magnitude makes the diode conduct. As a result, valve *I* is temporarily cut off and the current through valve *II* is doubled. The output pulse appears on the anode of valve *II*.

[5]) See e.g. E. J. van Barneveld, Fast counter circuits with decade scaler tubes, Philips tech. Rev. **16**, 360-370, 1954/55.

[6]) J. J. van Zolingen, Gevoelige triggerschakelingen, Ned. T. Natuurk. **23**, 42-55, 1957 (in Dutch).

presently explain, to make the diode conduct to such an extent that the loop gain becomes greater than unity. In valve I the current then drops to zero, in valve II it becomes twice as high as it was, resulting in a pulse of virtually constant amplitude at the anode of valve II.

We shall now examine the process in more detail. A negative pulse of, say, 20 mV on the grid of I reduces the anode current of this valve and thus increases its anode voltage. The diode D_1, which was hitherto blocked, then conducts, consequently increasing the grid potential (and hence the cathode potential) of valve II via the coupling capacitor C. Since the cathodes have a common cathode load, the cathode potential of valve I also increases, causing a further drop in the current through this valve. This process, then, leads to a very rapid regenerative cut-off of valve I. The cathode inductance L_k tends to keep the current flowing through it at a constant value, so that any increase in the current of valve II can occur only at the cost of the current through valve I. This effect helps to speed up the change-over. The rise in the current through II produces a drop in the anode potential, resulting in a negative voltage pulse at the output.

Since the circuit can be designed to operate reliably with input pulses of only 12 mV, we call it a "millivolt discriminator".

The favourable characteristics of the circuit are due in the first place to the high transconductance at the operating point of both conducting valves, and in the second place to the fact that the diode requires a smaller voltage excursion (*fig. 3*) to make it conduct than a triode or pentode; moreover it is not the input pulse itself that makes the diode conduct, but the pulse amplified by valve I.

In the circuit described the diode passes no current in the stable state. Discriminator circuits have been described elsewhere [7]) in which use is also made of a diode, but in the stable state this passes a current that depends on the discrimination level. The biasing of the valves is related to that of the diode, so that in these cases the point of maximum transconductance cannot usually be reached.

Fig. 4 shows the circuit diagram with some details added:

1) The anode load of valve I is an inductance L_a of 10 μH in parallel with a damping network consisting of a diode D_2 and a series resistor R_d. The resonant circuit formed by L_a with its stray capacitance is damped by R_d-D_2 such that, when valve I is cut off, a positive pulse with a single negative undershoot appears at the anode, as drawn in fig. 4. This waveform minimizes the accumulation of charge of the coupling capacitor C.

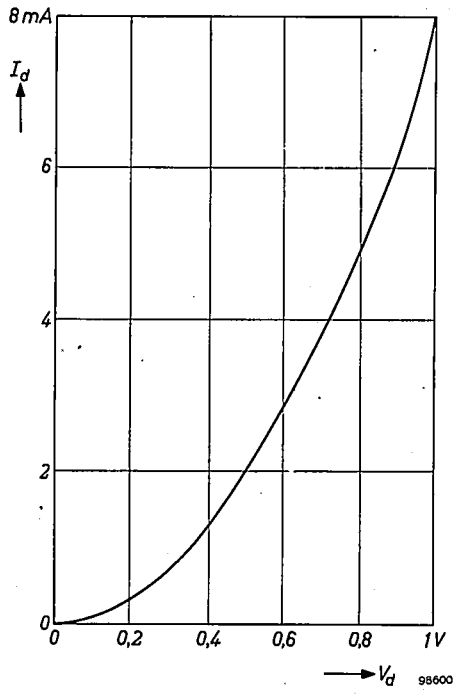


Fig. 3. D.C. characteristic (current I_d as a function of voltage V_d in the forward direction) of germanium diode type OA 85 (D_1 in figs. 2 and 4).

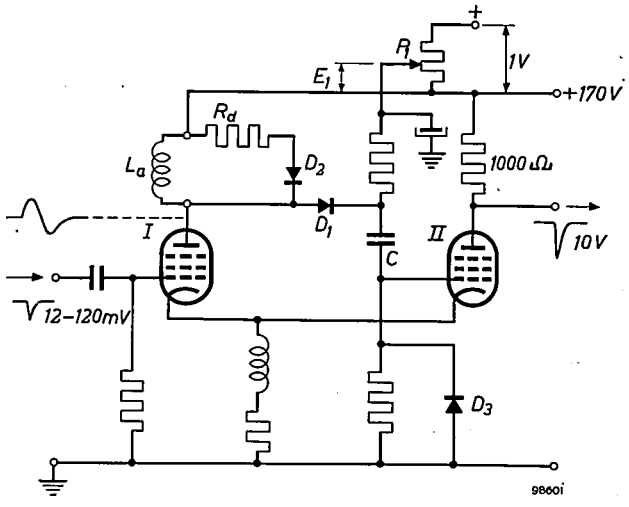


Fig. 4. Circuit diagram of millivolt discriminator, showing more details than fig. 2. Valves I and II are E 180 F pentodes, D_1 is a germanium diode, type OA 85. The anode load of valve I consists of a coil L_a (inductance 10 μH) shunted by a damping resistor R_d in series with a diode D_2. An input pulse of sufficient magnitude (≧ 12 mV) causes the anode voltage of I to rise. D_3 is a clamping diode through which any excess charge on the coupling capacitor C has a rapid path to earth. Potentiometer P_1 is used for adjusting the biasing voltage E_1 (0-1 V) which blocks D_1; in this way the discrimination threshold can be varied from 12 to 120 mV. The output pulse has an amplitude of 10 V across an anode load of 1000 ohms.

[7]) K. Kandah, A sensitive pulse trigger circuit with a stable threshold, Proc. Instn. Electr. Engrs. **101-II**, 239-247, 1954. S. Barabaschi, C. Cottini and E. Gatti, High sensitivity and accuracy pulse trigger circuit, Nuovo Cimento Ser. X, 2, 1042-1051, 1955.

2) Any excess charge on C can rapidly leak away through the clamping diode D_3.
3) Since the D.C. voltage drop across L_a is very small, the diode D_1 is blocked in the stable state by the biasing voltage E_1, the polarity of which is opposite to that in fig. 2. E_1 can be varied with the potentiometer P_1, making the discrimination threshold variable from 12 to 120 mV.

The measures mentioned under (1) and (2) give the circuit the desired virtues of a discrimination threshold and a dead time that depend to an almost negligible extent on the count rate. Measurements made with statistically distributed pulses have shown that the circuit possesses a dead time of 1.8×10^{-7} sec $\pm$ 20%. This value remains constant, within the limits of observation errors, up to count rates of 2.5×10^6 pulses per second.

Fig. 5a shows the biasing voltage E_i measured as a function of the amplitude V_i of the input pulses at which the circuit is just triggered. Except at the lowest point, the deviation from linearity is no more than 2% of the local value. The measurement was made with pulses of the shape illustrated in fig. 5*b*: total pulse-duration 10^{-7} sec, rise time about 1×10^{-8} sec, decay time 4×10^{-8} sec (rise time is defined as the time taken by the pulse to rise from 10% to 90% of its peak amplitude, and decay time as the time taken by the pulse to drop from 90% to 10% of its peak amplitude). The circuit has a large bandwidth, enabling it to deal with pulses of extremely short duration, as follows from the fact that the sensitivity was found to be only 25% lower for pulses of the shape shown in fig. 5*c* (total duration 1.8×10^{-8} sec, rise and decay times 0.75×10^{-8} sec).

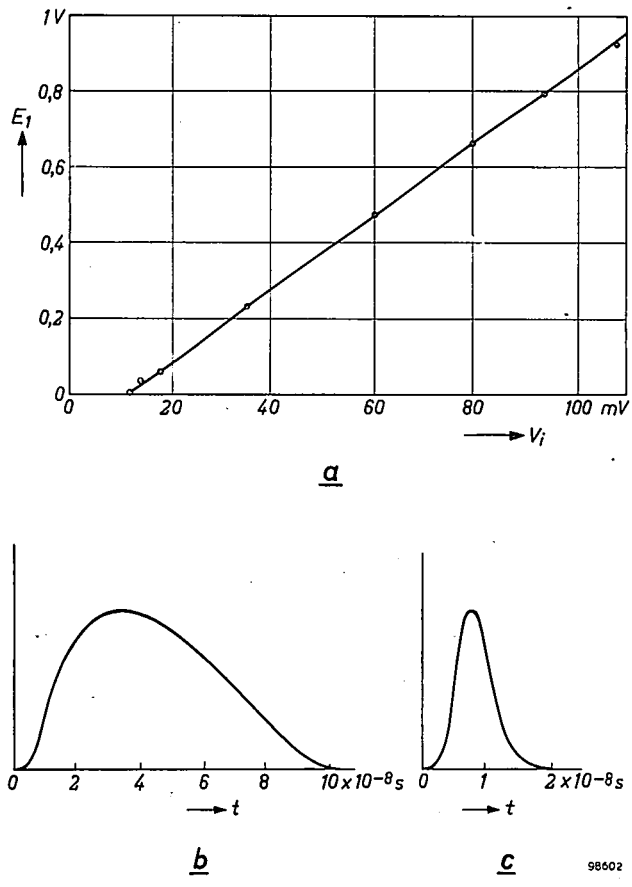


Fig. 5. *a*) Biasing voltage E_1 measured as a function of the threshold amplitude V_i which the input pulses must have in order to trigger the circuit in fig. 4.
b) Shape of the pulses used for the measurement of (*a*).
c) Pulse for which the circuit is only 25% less sensitive.

The large bandwidth and sensitivity are mainly due to the use of high-slope pentodes (type E 180 F) and a low-capacitance diode (germanium diode OA 85). Temperature variations of 18 °C were found to have scarcely any effect on the discrimination threshold.

The output pulses have an amplitude of 10 V across a resistance of 1000 ohms.

Stability of the millivolt discriminator

The stability of the millivolt discriminator was tested by applying to the circuit statistically distributed pulses from a scintillation counter at the mean rate of approximately 2×10^6 per second. The counter consisted of an anthracene crystal mounted on a photomultiplier tube and excited by gamma radiation from a ^{60}Co source. The pulses from the photomultiplier were applied directly to the millivolt discriminator, biased to a discrimination threshold of 30 mV. Repeated measurements were made throughout one week, during which the discrimination threshold referred to corresponded to the pulse height at the steep side of the Compton continuum in the energy spectrum of the electrons liberated by the gamma rays [8]). It was found that the discrimination threshold was stable to within 3%, and this small fluctuation was certainly not entirely due to the discriminator but must have been partly attributable to the photomultiplier, whose gain would not have been constant owing to the temperature variations that occurred.

The count-rate meters

For converting pulses of standard amplitude into a direct current proportional to the neutron flux Φ, or to log Φ, an integrating element is required. For this purpose we use a *pump circuit* (or a series of pump circuits). The waveform of the output voltage of the discriminator discussed above is

[8]) See e.g. J. A. W. van der Does de Bye, The scintillation counter, Philips tech. Rev. **20**, 209-219, 1958/59 (No. 8).

not suitable, however, to be applied to a pump circuit, and it was therefore necessary to include a pulse shaper between these two elements. We shall return to the pulse shaper after we have discussed the pump circuits, having then made clear the requirements to which the pulses must conform.

The basic form of the pump circuit is shown in *fig. 6*. It consists of a "reservoir capacitor" C_1, two diodes D_4 and D_5, and a "buffer capacitor" C_2

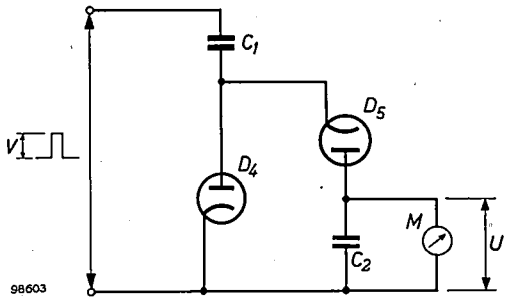


Fig. 6. Principle of pump circuit. C_1 reservoir capacitor. C_2 buffer capacitor. D_4, D_5 diodes; M milliammeter.

across which a meter M is shunted having a resistance R. Positive rectangular pulses of amplitude V are applied to the input at the rate of n per second. Each pulse, provided its duration is long enough, charges C_1 to the voltage V via D_4. At the end of the pulse a charge of magnitude $C_1(V-U)$, where U is the voltage across C_2, passes from C_1 to C_2 via D_5. If the time constant C_2R is large with respect to the period $1/n$ of the pulses, we can regard U as a constant (ripple-free) voltage. The charge supplied per second to C_2 via D_5 is thus $nC_1(V-U)$, and this current is equal in the steady state to the current U/R flowing through the meter. This equality leads to the following expression for the output voltage U:

$$U = \frac{nC_1R}{1+nC_1R}V = \frac{n\tau}{1+n\tau}V, \quad \ldots \quad (2)$$

where τ is the time constant C_1R.

The linear count-rate meter

In order to measure the count rate n on a linear scale, we must have a voltage proportional to n. From (2) we see that the relation between U and n is not a linear one. It is linear to a good approximation however, provided that:

$$n\tau \ll 1, \quad \ldots \quad (3)$$

for in this case (2) becomes:

$$U \approx n\tau V \ (\ll V). \quad \ldots \quad (4)$$

If this condition is satisfied, and if also the amplitude V of the input pulses is constant, we can read the neutron flux on a linear scale on the (moving-coil) meter.

The range of count rates in which the linear relationship is valid to a sufficient approximation can be appreciably extended by adding a negative-feedback circuit including a D.C. amplifier, as shown in *fig. 7*. The linearity condition of equation (3) can now be satisfied without the voltages U' and U'' (fig. 7) being small with respect to V. We now have:

$$U = U' + U''$$

and, if the gain of the amplifier is A:

$$U'' = -AU,$$

$$U' = (A+1)U.$$

Thus, U' and U'' have virtually the same magnitude but opposite polarity. At full deflection of the meter, the voltage U required from the integrating circuit is smaller by a factor A than in the circuit of fig. 6, which means that the desired linearity is obtained up to very much higher count rates. The amplitude of the deflection is virtually unaffected by A, provided that $A \gg 1$, e.g. ≈ 500; this is evident since U'', to which the deflection is proportional, is $-U'A/(A+1) \approx -U'$, hence practically independent of A.

The linear count-rate meter in the KEMA reactor is provided with a feedback circuit of this kind.

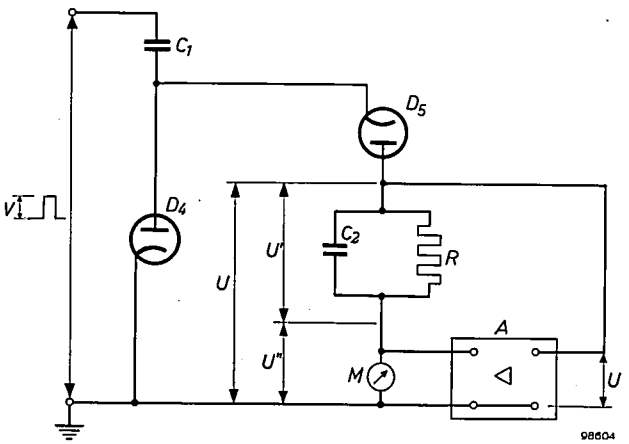


Fig. 7. Pump circuit with negative feedback via a D.C. amplifier A, which increases the range of count rates n over which the deflection of the moving-coil meter M varies linearly with n.

The logarithmic count-rate meter

To obtain a count-rate meter with a logarithmic scale, we make use of a number of pump circuits whose time constants form a geometric progression,

as described by Cooke-Yarborough and Pulsford [9][10]). If, for example, we take three pump circuits whose time constants are τ_0, $10\tau_0$ and $100\tau_0$, respectively, feed the same pulses to these circuits and add the three output voltages together, we find that the total output voltage increases more or less linearly with the logarithm of the count rate n over a range of about two decades (*fig. 8*). (For reasons of circuitry it is easier to add the output currents than the output voltages; fig. 8 accordingly refers to the currents.) The use of more pump circuits widens the range in which the total current increases approximately linearly with log n.

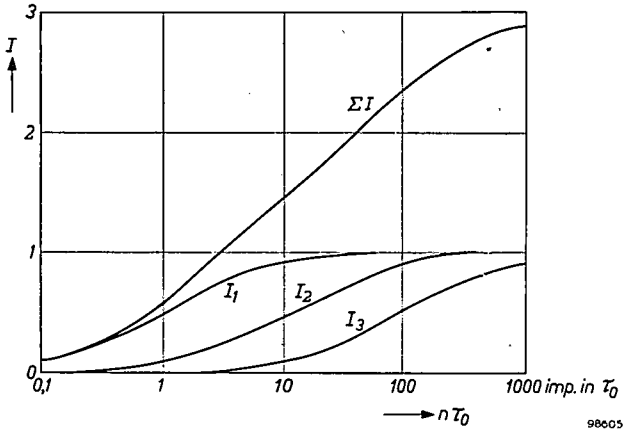


Fig. 8. Curves I_1, I_2 and I_3: output current of pump circuits with time constants τ_0, $10\tau_0$ and $100\tau_0$, as a function of the number of pulses $n\tau_0$ in the time τ_0. The sum ΣI of I_1, I_2 and I_3 varies nearly linearly with log n over about two decades. This range is widened as more pump circuits are connected in parallel.

If τ_r is the time constant of the r^{th} pump circuit, and if the count rate changes from n to an, then the change δU_r in the output voltage of this pump circuit is

$$\delta U_r = \left(\frac{an\tau_r}{1+an\tau_r} - \frac{n\tau_r}{1+n\tau_r}\right) V,$$

where V is again the amplitude of the pulses.

We now consider an infinitely large number of parallel pump circuits [9]), whose output voltages are added together and whose time constants form a geometric progression of ratio k, so that for the r^{th} circuit:

$$\tau_r = \tau_0 k^r.$$

The change δU_0 in the total output voltage is then given by:

$$\delta U_0 = \sum_{r=-\infty}^{r=+\infty} \delta U_r = \sum_{-\infty}^{+\infty} \left(\frac{an\tau_0}{k^{-r}+an\tau_0} - \frac{n\tau_0}{k^{-r}+n\tau_0}\right) V. \quad (5)$$

The result of this summation, writing $\log k = \lambda$, is:

$$\delta U_0 = \left[\frac{\log a}{\lambda} - \frac{4\pi}{\lambda} \exp\left(\frac{-4\pi^2}{\lambda}\right) \sin\left(\frac{2\pi}{\lambda}\log a\right)\right] V. \quad . \quad . \quad (6)$$

The first term in the right-hand side of (6) yields the required logarithmic relation between δU_0 and a. The second term represents a deviation from the log law but is relatively small (of the order of a fraction of 1%) and recurs periodically from decade to decade (i.e. with the period $\log a$, that is by the same amount for a, ka, k^2a, k^3a, etc.).

Of course, in practice the number of circuits that can be connected in parallel is limited. Deviations will therefore occur from the logarithmic relation both at low and high count rates. It is found, however, that the deviations are appreciably reduced if the pulses fed to the two last pump circuits are given a somewhat larger amplitude, e.g. 15% larger, than the pulses for the other circuits. In this way it has proved possible with six pump circuits to cover a range of five decades, and with eight pump circuits a range of $6\frac{1}{2}$ decades. *Fig. 9* gives the circuit for the latter case, together with the values of the resistances and capacitances. The scale is logarithmic from $n = 1$ to $n = 3\times10^6$ pulses per second, as the calibration curve in *fig. 10* shows. By virtue of these $6\frac{1}{2}$ decades one can readily observe the variation with time of the neutron flux from source level up to the range where the less sensitive neutron-flux meter, which measures the current of an ionization chamber, begins to give a reading (see the following article, IIIB).

In fig. 9 three pump circuits contain a resistor in parallel with the buffer capacitor C_2. This makes it possible to use reservoir capacitors C_1 of greater capacitance than the few picofarads required by equation (6), so reducing the effect of incidental changes in the stray capacitances of the wiring.

The capacitance C_1 in these three pump circuits is made larger than in the others in order to compensate to some extent for the statistically distributed pulses that are not counted, owing e.g. to the finite resolution of the discriminator or of the pulse shaper. In this way the calibration characteristic for statistically distributed pulses is linearized. For periodic pulses, however, the characteristic is somewhat curved at very high repetition frequencies.

The pulse shaper

For linear as well as for logarithmic count-rate meters the pulses are required to be of constant amplitude, irrespective of the characteristics of the valves used in the circuits that produce the pulses, and also at the highest count rates n that occur.

A further requirement in the logarithmic counter is that the pulse duration should increase as n decreases. The smaller n becomes, the more pump

[9]) E. H. Cooke-Yarborough and E. W. Pulsford, An accurate logarithmic counting-rate meter covering a wide range, Proc. Instn. Electr. Engrs. **98-II**, 196-203, 1951.

[10]) For other methods of obtaining a logarithmic scale covering numerous decades, see: F. E. L. ten Haaf, G. Klein and F. J. Schijff, Monitoring, control and safety equipment for a nuclear reactor of the swimming-pool type, II. Further description of certain component units, Philips tech. Rev. **19**, 273-285, 1957/58.

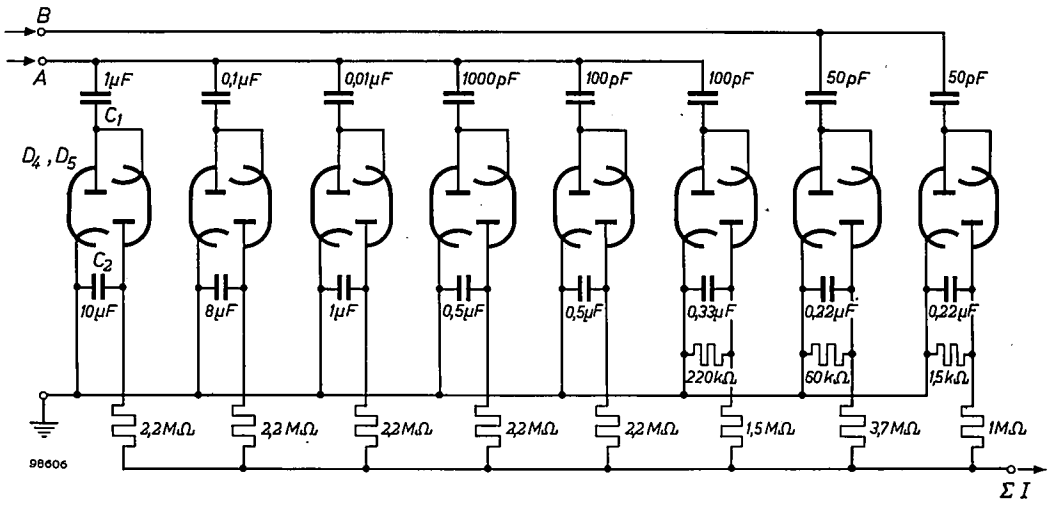


Fig. 9. Parallel arrangement of eight pump circuits. Meaning of C_1, C_2, D_4 and D_5 as in fig. 6. The pulses applied to terminal B are approximately 15% larger than the pulses applied to terminal A. Bottom right is the terminal from which the total output current ΣI is taken. The valves are double diodes, type EAA 91.

circuits in which C_1 has a high capacitance (the circuits on the left in fig. 9) will contribute to the output current, and hence the longer the time needed for these capacitors to be fully charged.

This requirement is certainly not fulfilled by the pulses from the millivolt discriminator discussed above (fig. 4). For this reason, use is made of a bi-stable flip-flop circuit that produces rectangular pulses, the average width of which increases automatically the fewer pulses are supplied to the circuit per second.

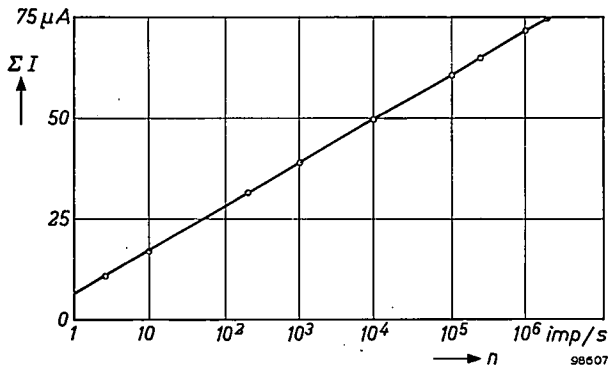


Fig. 10. The output current ΣI of the circuit in fig. 9 as a function of the count rate n. The relation between ΣI and log n is virtually linear over $6\frac{1}{2}$ decades (deviation from linearity nowhere $>10\%$ of the local value).

A flip-flop circuit described by Cooke-Yarborough and Pulsford [9]) delivers pulses that fully meet the requirements as long as n does not exceed a value of approximately 10^5 pulses per second. In our case, however, values of n occur that are more than a factor of 10 higher, and these cannot be handled by the circuit referred to. For this reason an entirely new flip-flop circuit has been designed, capable of particularly rapid operation. This is used in conjunction with an amplitude stabilizer, which in its turn is followed by an output stage; the arrangement is shown in the block diagram in *fig. 11*.

A brief description of these three components is given below.

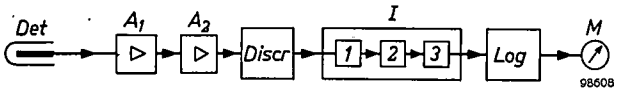


Fig. 11. Between the millivolt discriminator *Discr* and the count-rate meter *Log* a pulse shaper *I* is interposed, consisting of three sections: a fast flip-flop *1*, an amplitude stabilizer *2* and an output stage *3*.

Fast-response flip-flop circuit

The circuit diagram of the pulse shaper, somewhat simplified, is shown in *fig. 12*, section *1* being the flip-flop circuit. The latter is required to have an extremely short dead time (less than 10^{-7} sec) and also to function reliably at count rates as high as several millions per second.

The dead time is shorted by reducing all the time constants involved, in this case the time constant of the anode impedances and that of the coupling networks between the valves. A limit is set to this, however, owing to the fact that if the values of resistance or capacitance are too small, the valves will no longer cut each other off.

Two routing diodes D_{I} and D_{II} are used at the input. These conduct the negative driving pulse to the most sensitive place in the circuit, that is to the grid of the conducting valve. If, however, the count rate exceeds a specific value, the coupling capacitor C_c no longer has sufficient time to dis-

charge. As a result the potential of point Q increases, thereby blocking the routing diodes. This drawback is overcome by clamping the potential of Q with the aid of the diode D_c at a value that can be varied with the potentiometer P_2. This makes it possible to go up to much greater count rates. Fitted with a double triode type E 92 CC and anode resistances of 3300 ohms, the circuit proved capable of handling reliably 15×10^6 regularly-spaced pulses per second [11]).

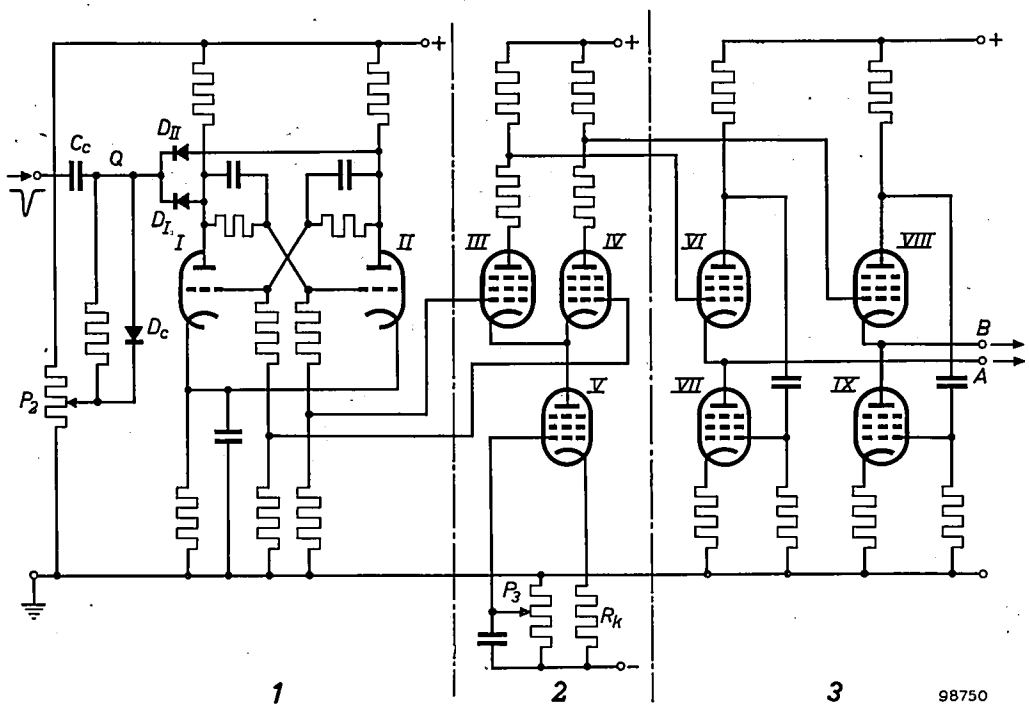


Fig. 12. Circuit diagram of pulse shaper I of fig. 11 (somewhat simplified).
Section 1: fast flip-flop. *I-II* double triode E 92 CC; one half conducts while the other is cut off. D_I, D_{II} routing diodes which conduct the incoming negative pulses to the grid of the conducting half of the double triode. (Suppose that I is cut off and II conducting. In that case I has a higher anode voltage than II, so that diode D_I is open and diode D_{II} blocked. A negative pulse thus arrives on the anode of I, i.e. on the grid of II via the associated coupling capacitor.) D_c clamping diode which fixes the potential of point Q at a value that can be varied with potentiometer P_2; this prevents blocking of the routing diodes at very high count rates.
Section 2: amplitude stabilizer. The pulses from the flip-flop circuit drive the pentodes *III* and *IV* (type E 180 F) such that one passes current while the other does not. The pentode *V* (type EF 80) with its cathode resistor R_k acts as a constant current source, the magnitude of the current being adjustable with potentiometer P_3. As a result the voltage pulses tapped from the anode resistors are of constant amplitude.
Section 3: output stage, consisting of two White cathode followers (pentodes *VI-VII* and *VIII-IX*, all type EL 83). This cathode-follower arrangement has the merit of a very low effective input capacitance and output impedance. The output pulses from *VIII-IX* are 15% greater in amplitude than those from *VI-VII* (cf. fig. 9).

The amplitude stabilizer

Section 2 of fig. 12 represents the amplitude stabilizer. The square-wave output of the flip-flop *1* alternately cuts off one of the two pentodes, type E 180 F. The current through the other one is determined entirely by a third pentode (an EF 80) in the common cathode lead of the two E 180 F valves. This third pentode functions as a source of constant current (partly by virtue of the cathode resistor R_k, which is not decoupled and which therefore introduces negative feedback); the magnitude of the current can be varied with potentiometer P_3. Since the current flowing through one of the two E 180 F pentodes is thus held constant, the voltage drop it causes in the anode resistance is also constant. This means that the output pulses have a constant amplitude.

Output stage of pulse shaper

If the pump circuits were to be connected directly to the amplitude stabilizer, they would constitute such a heavy capacitative load that the pulse edges would no longer be steep enough for the pulses in a rapid succession to retain a flat top and constant

[11]) Using a double triode E 88 CC (which has a higher transconductance than type E 92 CC) in conjunction with anode resistances of 1200 ohms and suitably adapted coupling networks, it has proved possible to reach count rates as high as 30×10^6 pulses per second for regularly-spaced pulses.

amplitude. For this reason the amplitude stabilizer must be followed by an output stage which on the one hand constitutes a very low capacitative load on the stabilizer and on the other has a low output impedance.

In both respects the White cathode-follower arrangement using two pentodes in series (see section *3* of fig. 12) is particularly efficient; EL 83 pentodes are used in order to obtain a square-wave output of 30 V amplitude. The effective input capacitance C_i is in this case:

$$C_i = (1-A)C,$$

where C is the "geometric" input capacitance and A the amplification. Since the latter is somewhat less than unity with a cathode follower, C_i is much smaller than C. The output impedance Z_0 is:

$$Z_0 = \frac{R_i\,(R_i + R_a)}{(S^2R_i^2 + SR_i + 1)R_a + (SR_i + 2)R_i},$$

where R_i is the internal resistance and S the slope (transconductance) of the valves, and R_a the anode load. If we substitute for R_i and S in this formula the values applicable to the EL 83, we find that R_a can be so chosen that Z_0 assumes a very low value, of the order of ten ohms. Such a value is quite small enough for feeding a series of pump circuits, even where pulses are involved with rise and decay times of about 4×10^{-8} sec.

Performance of the logarithmic counting channel

It will be evident that with the pulse shaper and pump circuits described above it is possible to build both linear and logarithmic count-rate meters capable of very high count rates. Calibrations have confirmed this. As regards the logarithmic instruments, fed for example by pulses from a scintillation counter, we have already seen in fig. 10 that the linear relation between the meter deflection and log n is still valid at $n = 3\times10^6$ pulses per second. The deviations from linearity in the whole range of $6\frac{1}{2}$ decades nowhere exceed 10% of the local value.

In conjunction with the millivolt discriminator discussed earlier, this logarithmic count-rate meter lends itself particularly well for measuring the neutron flux in a nuclear reactor. Photographs of the instruments as used in the KEMA reactor can be seen in *fig. 13a* and *b*.

The neutron detector itself must obviously be capable of delivering pulses at the high rates mentioned. The detector containing ^{10}B cannot do this, but the fission chamber can. The drawback that a gain of about 10^5 is required between this counter tube and a discriminator of the conventional type is not applicable in the apparatus described above; in this case the gain (in fig. 1, from A_1 and A_2 together) need only be 100.

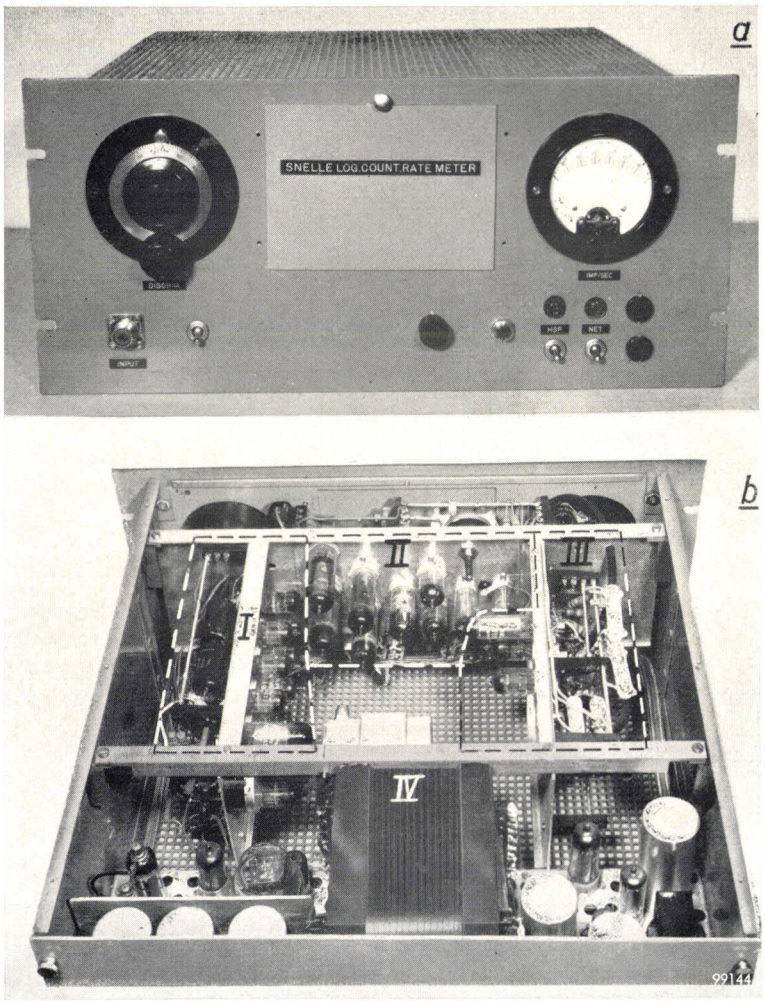


Fig. 13. Pulse-counting channel with logarithmic scale as developed and built in the KEMA laboratory. *a*) Front view; left, the potentiometer knob for setting the discrimination threshold; right, the meter with a logarithmic scale covering $6\frac{1}{2}$ decades. *b*) Interior view, showing the pump circuits *I*, the pulse shaper *II*, the discriminator *III* and the power pack *IV*.

Summary. The article describes the pulse-counting channels used in conjunction with a neutron detector for monitoring the neutron flux in the KEMA subcritical suspension reactor at Arnhem. Various components of the system were developed

and built in the KEMA laboratory, notably the "millivolt discriminator" which operates reliably with pulses from 12 mV at count rates of several million pulses per second, and an improved fast-response flip-flop in the pulse shaper, capable of handling up to 15 million pulses per second. The linear integrator consists of a "pump circuit", and the logarithmic integrator of eight pump circuits in parallel. In the latter a logarithmic relation exists between the meter deflection and the count rate n, extending from $n = 1$ to $n = 3\times10^6$ pulses per second ($6\frac{1}{2}$ decades). This property makes the logarithmic count-rate meter very suitable for monitoring the increase in neutron flux when a nuclear reactor is started up.

IIIB. THE MONITORING OF HIGH NEUTRON FLUX WITH THE AID OF AN ELECTROMETER

by M. van TOL.

621.039.564.2:621.039.524.46

In the case of the KEMA subcritical suspension reactor [1]), "high" neutron flux comprises flux values from about 2×10^3 to 2×10^6 neutrons per cm² per second, corresponding to a generated power of 1 mW to 1 W. To measure the flux in this range, the direct-current output of an ionization chamber filled with gaseous boron-trifluoride ($^{10}BF_3$) is measured [2]).

The design of the monitoring system was based on the following considerations:

1) At a neutron flux of 2000 neutrons per cm² per second, the ionization chamber delivers a current of 2×10^{-11} A.
2) This current — the smallest required to be measured — must produce a distinctly readable deflection; the latter is taken to be four scale divisions on a linear scale graduated up to 100.
3) For safety reasons the neutron-flux meter must remain in operation in the event of a mains failure.
4) For the same reasons the system must consist of two identical channels, each channel being fully independent of the other.
5) Each of the two channels must also serve to vary the position of a control rod via a servo-mechanism, thus keeping the neutron flux constant.
6) Both channels must be designed for connection to a recording instrument and a remotely positioned meter.
7) It must be possible to stop the reactor within 0.5 sec in any of the following contingencies:
   a) if the neutron flux exceeds a certain level,
   b) in the event of a mains failure,
   c) if the voltage on the ionization chambers falls off,
   d) if a defect occurs in the neutron-flux meters themselves.

Some of these points will be discussed at greater length below.

Electrometer with vibrating capacitor

For measuring the ionization current, which is of the order of 10^{-11} to 10^{-7} A, use is made of a vibrating-reed electrometer based on a principle which has been described in this journal [3]). The current passes through a high resistance R_1, and the direct voltage produced across it is converted into an alternating voltage by means of a vibrating capacitor. This is an air capacitor consisting of two parallel plates, one of which is stationary and the other kept in vibration. When the direct voltage is applied to the vibrating capacitor, an alternating voltage is produced between the plates which is proportional to the voltage across R_1 (cf. the operation of a condenser microphone). The alternating voltage is amplified and rectified, producing at the output a direct current that causes the deflection of a moving-coil meter. The deflection is a linear measure of the neutron flux. The vibrating capacitor is driven by a kind of loudspeaker movement powered by a valve oscillator; the frequency is 140 c/s. A block diagram is shown in *fig. 1*.

To prevent the sensitivity from being affected by mains voltage fluctuations and changes in valve characteristics, strong negative feedback is applied: a certain fraction β of the direct voltage output E_2 is added in opposite polarity to the input voltage E_1 across the high resistance R_1. The direct voltage converted into an alternating voltage is thus $E_1 - \beta E_2 = E_1'$. Taking the ratio $E_2/E_1' = A'$, the gain ultimately obtained, $E_2/E_1 = A$, is found to be:

$$A = \frac{E_2}{E_1} = \frac{A'}{1+\beta A'}.$$

[1]) J. J. Went, Instrumentation for a subcritical homogeneous suspension reactor, I. Reasons behind the choice of a homogeneous suspension reactor, Philips tech. Rev. **21**, 109-121, 1959/60 (No. 4/5).

[2]) See the reaction (1) in article IIIA of this series (p. 135).

[3]) J. van Hengel and W. J. Oosterkamp, A direct-reading dynamic electrometer, Philips tech. Rev. **10**, 338-346, 1948/49.

If $\beta A'$ (the loop gain) is large with respect to 1, we can write, to a good approximation:

$$A \approx \frac{1}{\beta},$$

so that the ultimate gain A is virtually independent of the gain A' of the valve circuit, which varies with the supply voltages and with the characteristics of the valves. In our case $\beta A' = 50$; with this value, a very constant value of A is obtained.

The values of β and A', separately, follow from the extreme values of E_1 and E_2. The lowest current required to be measured, 2×10^{-11} A, must produce a deflection of four scale divisions. The full deflection of 100 scale divisions therefore corresponds to a current of $25\times2\times10^{-11} = 5\times10^{-10}$ A. This produces across the resistor R_1, which is 200 MΩ, a voltage E_1 of 0.10 V. The output voltage E_2 required to cause full deflection of the meter is 2 V. The total amplification must therefore be $E_2/E_1 = 20$. For this purpose $^1/_{20}$ of the output voltage must be fed back to the input. Since we decided on $\beta A' = 50$, the nominal gain A' must be 1000.

Measuring ranges

As mentioned above, the high neutron-flux range covers three decades. Since the scale is linear, this range must be divided into separate ranges, in this case into six. The same constancy is required for all these ranges ($\beta A' = 50$) and in each case a value of $E_2 = 2$ V is required for full deflection.

The values referred to above, viz. $\beta = {}^1/_{20}$, $A' = 1000$ and $R_1 = 200$ MΩ, apply to the measuring range having the greatest sensitivity (position I): full deflection at 5×10^{-10} A. In positions II and III, β is greater by a factor of $\sqrt{10}$ and 10, respectively, and the sensitivity is therefore lower by the same factor; the currents that deflect the needle to the end of the scale are accordingly 1.6×10^{-9} A and 5×10^{-9} A, respectively. At the same time as β is increased, A' is reduced by the same factor, so that the product $\beta A'$ remains constant.

In positions IV, V and VI, β and A' have the same values as in positions I, II and III, but the value of resistance R_1 is reduced to $200/(10\sqrt{10})$ MΩ. In the range of lowest sensitivity (position VI) full deflection is therefore obtained with a current of $10\sqrt{10}\times5\times10^{-9} = 1.6\times10^{-7}$ A.

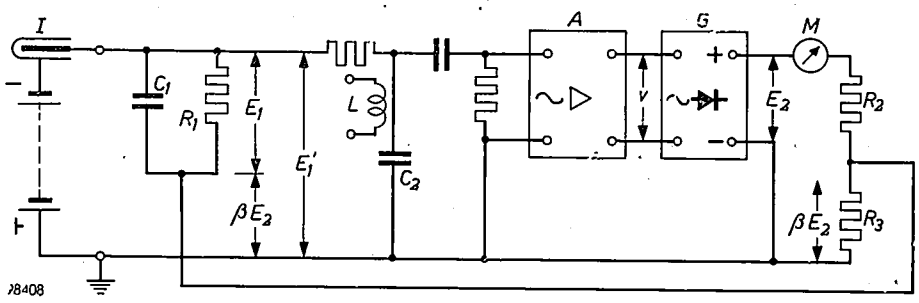


Fig. 1. Illustrating the principles of the monitoring system for high neutron flux. *I* ionization chamber. R_1 resistance of 200 MΩ (max.) with shunt capacitance C_1 for smoothing voltage variations caused by statistical fluctuations in the neutron flux. C_2 vibrating capacitor with driving coil *L*. *A* A.C. amplifier (with two EF 40 pentodes in cascade). *G* phase-sensitive rectifier (see fig. 4 and caption of the article quoted under [3]). *M* moving-coil meter. R_2-R_3 voltage divider from which the negative feedback voltage βE_2 is tapped.

Details of the voltage divider R_2-R_3 (fig. 1), from which the negative feedback voltage βE_2 is tapped, are shown in *fig. 2*. The figure also indicates the means of providing a signal required by the safety system (see article IV) as well as a connection for a recording instrument. The latter can be arranged to actuate a servomechanism, which varies the position of a control rod in such a way as to keep the neutron flux constant.

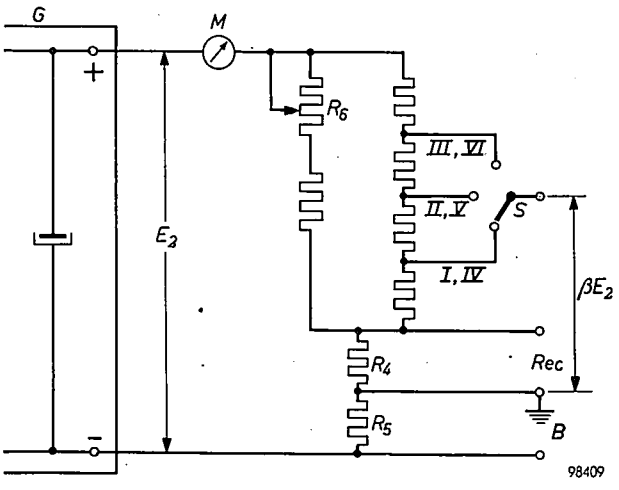


Fig. 2. Details of the voltage divider R_2-R_3 (fig. 1). *G* rectifier. *M* moving-coil meter. *S* measuring-range switch: $\beta = 1/20$ in positions I and IV, $\beta = \sqrt{10}/20$ in positions II and V, $\beta = 10/20$ in positions III and VI. Recording instrument *Rec* is connected in parallel with the low resistance R_4. The voltage across R_5 provides the signal actuating the safety system if certain limits are exceeded (see article IV); the safety system is connected at *B*. R_6 correction potentiometer.

Calibrating device

Neutron-flux meters are the chief instruments employed for controlling the operation of a nuclear reactor. If only for this reason it is desirable to be able to ascertain at any time whether these instruments are working properly. This is all the more important if, as in the KEMA reactor, the neutron-flux meters also supply the signal that puts the safety system into operation in the event of a dangerous situation arising.

The principle of the calibrating device is shown in simplified form in *fig. 3*. Fig. 3*a* shows the method of charging-up a capacitor C_3 of 50 μF which, during calibration (fig. 3*b* and *c*) acts as the source of anode current for an EF 40 pentode, whose control-grid and screen-grid potentials are constant. During the approximately 200 μA discharge, which takes place during the calibration (fig. 3*b* and *c*), the voltage on the capacitor obviously falls. For about 12 sec, however, this has no perceptible effect on the anode current, owing to the pentode property referred to, and partly owing to the negative feedback, effected via the cathode resistance R_8. During the calibrating proper (fig. 3*c*), a capacitative attenuator C_3-C_4 ensures that a certain fraction ($\frac{7}{50}\times10^{-6}$) of the said current flows for calibration purposes through R_1.

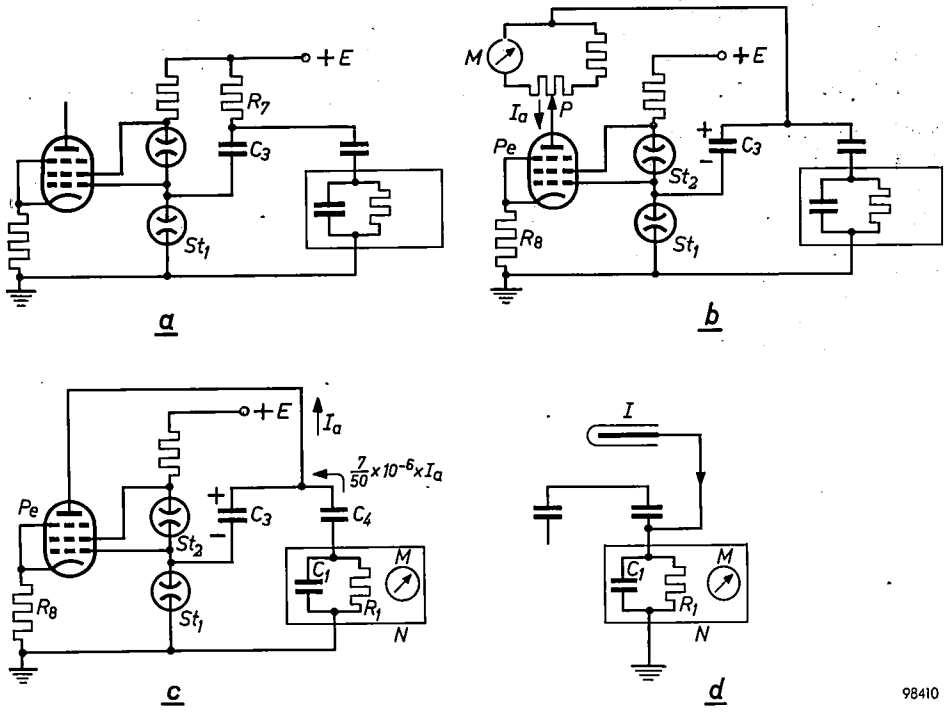


Fig. 3. Calibrating device for the neutron-flux monitoring system.
a) Via the resistor R_7 and the voltage-reference tube St_1 the capacitor C_3 (50 μF) is charged up from the direct-voltage source E.
b) The charged capacitor C_3 now discharges into the anode circuit of the pentode Pe (EF 40), whose control-grid and screen-grid potentials are kept constant by the voltage-reference tubes St_1 and St_2, and whose cathode resistance R_8 contributes to keeping the anode current I_a constant. This discharge current (approx. 200 μA) flows through the parallel circuit of the moving-coil meter M (now entirely disconnected from the neutron-flux monitor) and a shunt. The potentiometer P is preset such that M gives a certain standard deflection.
c) Checking the neutron-flux meter reading against the standard deflection obtained in (*b*). I_a now distributes itself over the capacitors C_3 (50 μF) and C_4 (7 pF) of a capacitative attenuator. The current through C_4 is $\frac{7}{50}\times10^{-6}\times I_a$ and flows through the resistor R_1 (200 MΩ) in the flux monitor N; the moving-coil meter M must now give the same reading as in (*b*).
d) The flux monitor in normal operation, connected to the ionization chamber I. The calibrating device is out of operation.

The neutron-flux meter described is therefore equipped with a calibrating device, with which a known current can be applied for a short time to the 200 MΩ resistor R_1. This current is obtained by making use of the property of a pentode that, when the control-grid and screen-grid potentials are kept constant, the anode current remains, for a wide range of potentials, virtually independent of the anode voltage.

The calibration procedure is as follows. After capacitor C_3 has been charged up as in fig. 3*a*, the system is switched to

the circuit shown in fig. 3*b*. C_3 now discharges with a current I_a of about 200 μA, kept constant by the pentode. Part of this current flows through the moving-coil meter M (fig. 1), which for the moment is entirely disconnected from the neutron-flux meter. A preset shunt (potentiometer P) ensures that the meter gives a certain standard deflection. Next, after recharging as in (*a*), the system is switched over to the circuit in fig. 3*c*. The meter (with unchanged shunt) is now connected in the normal way in the output circuit of the neutron-flux meter, and the constant current I_a distributes itself over the capacitor C_3 (= 50 μF) and C_4 (= 7 pF). The portion now flowing via C_4 amounts to $^7/_{50} \times 10^{-6} \times I_a$, and this is conducted through the input resistor of the neutron-flux meter and produces a reading on meter M. The ratio $C_3 : C_4$ of the capacitative attenuator is chosen such that the deflection, provided there is nothing wrong with the neutron-flux meter, is the same as that in the case of fig. 3*b*. Finally, fig. 3*d* shows the normal flux-monitoring situation, with the calibrating system switched off.

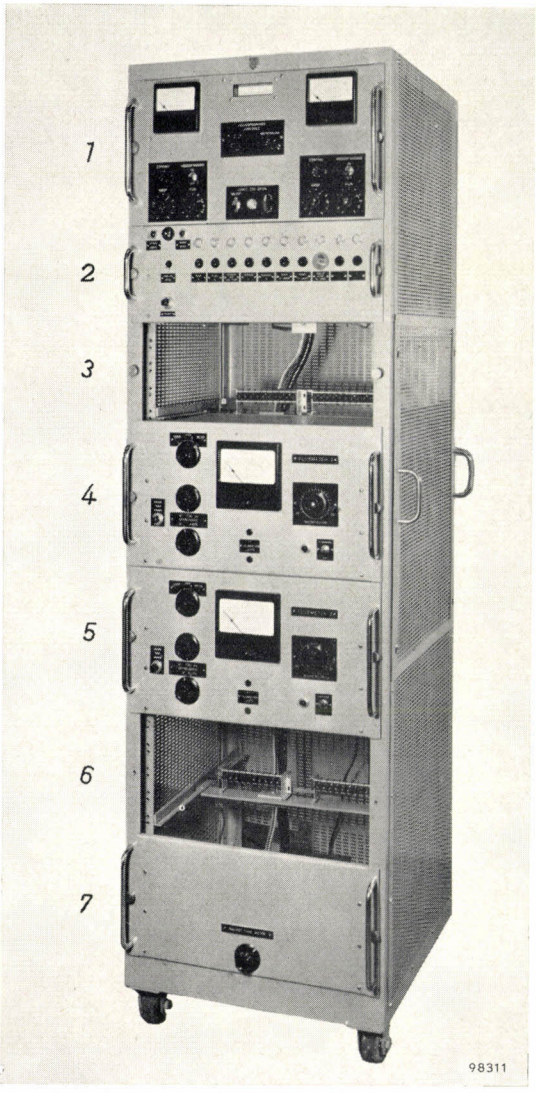


Fig. 4. The monitoring, control and safety equipment built by Philips for the KEMA subcritical suspension reactor. The units *1* to *7* contain the following:
Unit *1*: High-tension rectifiers and comparator circuit (with light-relay) for the potentials on the ionization chambers (see article IV of this series).
Unit *2*: Safety circuits, and pilot lamps that indicate which safety circuit has first operated (see article IV).
Unit *3*: Space for control panel of the temperature control system (see article II).
Units *4* and *5*: Identical neutron-flux monitoring channels based on electrometer system.
Unit *6*: Space reserved for future equipment.
Unit *7*: Unit of the temperature-protective system (see article IV).

Varying contact potentials may cause the meter M to deflect even though no current is applied to the neutron-flux meter. This error can be corrected with a potentiometer which supplies a small variable direct voltage in series with the vibrating capacitor.

Power supply

Each of the two identical flux-monotoring channels is powered by its own 6 V battery; they are thus entirely independent of the mains and independent of each other. Heater currents are taken directly from each battery, which also supply the current for D.C. motor generators that produce the H.T. voltages of 300 V.

The ionization chambers require a direct voltage of about 2500 V. This is obtained in a manner similar to that used for generating the E.H.T. for the picture tube in a television set [4]); the high-frequency output of an oscillator is stepped up by a transformer, rectified and smoothed to a direct voltage of 2500 V. Part of this voltage is compared with a reference voltage and the deviation from the latter is used for controlling the oscillator. In this way the E.H.T. is stabilized. The oscillator is fed from the battery and the motor generator.

Fig. 4 shows the actual equipment. Apart from the neutron-flux monitoring channels discussed above, it also contains the safety circuits, which are dealt with in the following article.

[4]) See e.g. Philips tech. Rev. **10**, 125, 1948/49, or **14**, 21, 1952/53.

Summary. Description of the monitoring system used in the KEMA subcritical suspension reactor for measuring the neutron flux in the range 2×10^3 to 2×10^6 neutrons per cm² per second. For this purpose an ionization chamber and vibrating-capacitor electrometer are used. The electrometer has strong negative feedback (loop gain 50); the meter has six measuring ranges and a linear scale. The minimum and maximum currents that produce full deflection are 5×10^{-10} and 1.6×10^{-7} A, respectively. A calibrating device is incorporated for checking the operation of the instruments at any time. The monitoring system consists of two identical channels; each channel is powered from its own battery.

IV. THE SAFETY CIRCUITS

by F. J. SCHIJFF.

621.039.524.46:621.039.587

Protective system with safety rods

Nuclear reactors must be safeguarded by a protective system against the possibly disastrous consequences of faults in operation or defects in the apparatus. If, for example, the neutron flux in the reactor should become excessive or increase too rapidly, if the temperature should become too high, or a neutron-flux meter fail to give a reading, or any other abnormality occur, the reactor must be automatically stopped. As a rule this is done by causing rods of neutron-absorbent material (safety rods) to drop quickly into the reactor, thereby discontinuing or preventing the occurrence of the chain reaction. A conventional arrangement is to suspend the safety rods from electromagnets (*fig. 1a*); when the current energizing the magnets is interrupted, the safety rods drop into the reactor. The magnet currents are interrupted by means of relays, which respond to the abnormalities against which the reactor is to be safeguarded. A protective system of this type has already been described in this journal [1]).

The KEMA subcritical suspension reactor [2]) is also safeguarded in this manner. In this case three safety rods of boron carbide are used, one of which serves as control rod. These rods do not fall into the reactor vessel but into the neutron reflector surrounding it; see fig. 5 of article I of this series.

Both for accurately measuring the various quantities (neutron flux, temperature, etc.) and for protective purposes, detectors are required that convert the quantity concerned into an electrical signal. In principle, the same detectors can serve both purposes. It is preferable, however, for a safety system to have its own detectors, since those used for accurate measurements are an integral part of a complicated apparatus and are required to be extremely sensitive. These conditions are not always compatible with the primary requirement of a protective system which is to ensure *maximum safety* (certain and prompt intervention when necessary) and reasonable *reliability* (failures that give rise to spurious interventions should be rare). The very

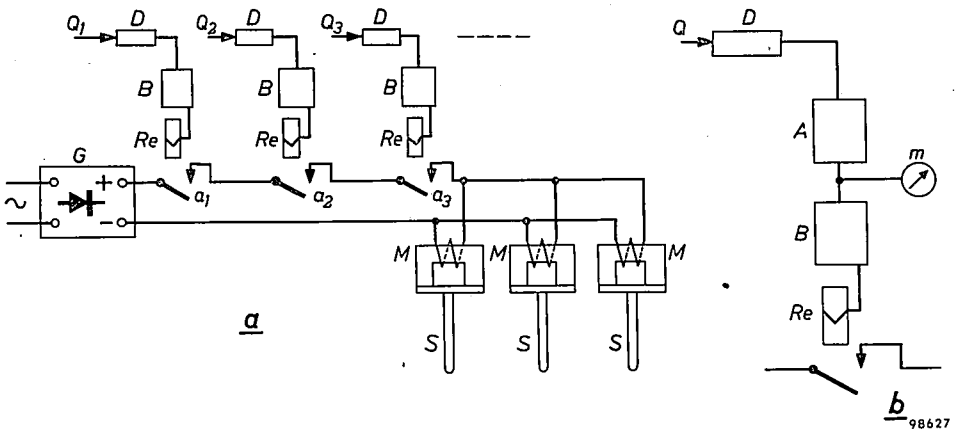


Fig. 1. *a*) Principle of a safety circuit for a nuclear reactor. The detectors D for the quantities Q_1, Q_2, Q_3, ... (neutron flux, temperature, etc.) supply a signal to the circuits B, each of which energizes a relay Re. Via the relay contacts a_1, a_2, a_3, ... the rectifier G energizes the electromagnets M from which the safety rods S are suspended. If one of the quantities Q reaches a critical value, the associated relay cuts out, the current through the electromagnets is interrupted and the rods S drop, thereby preventing any further increase in the neutron flux.
b) In the KEMA reactor there is no space for separate detectors in the safety systems. The circuits B obtain their signal from the measuring channels A (with detector D and meter m); A may be e.g. a neutron-flux meter, a period meter, a temperature meter, a flow meter, etc.

[1]) F. E. L. ten Haaf, G. Klein and F. J. Schijff, Monitoring, control and safety equipment for a nuclear reactor of the swimming-pool type, II. Further description of certain components units, Philips tech. Rev. **19**, 273-285, 1957/58, in particular pages 281-285.

[2]) J. J. Went, Instrumentation for a subcritical homogeneous suspension reactor, I. Reasons behind the choice of a homogeneous suspension reactor, Philips tech. Rev. **21**, 109-121, 1959/60 (No. 4/5).

compact design of the KEMA reactor, however, does not permit the use of separate detectors in the safety system; the same detectors, and to some extent the same amplifiers, therefore had to serve both for measurements and protection (fig. 1*b*). The difficulty was increased by the fact that measurements were to be made on this reactor at very low levels of neutron flux, which called for an extremely sensitive, and therefore delicate, measuring system.

The presence of complicated instruments in the protective system essentially entails a lower degree of both safety and reliability. Now the experimental nature of the KEMA reactor makes it possible to tolerate more frequent unnecessary scrams — i.e. a lower degree of safety-equipment reliability — than would normally be permissible. The following measures could therefore be taken to improve the safety factor, although with some sacrifice of reliability.

1) The principal measuring channels — likewise safety channels — were duplicated.
2) Upper-limit as well as lower-limit protection is provided: the safety circuit coupled to the measuring channel intervenes not only when the parameter concerned exceeds a certain value, but also when the output signal of the measuring channel is too small (which might indicate a defect).

In this way a safeguard is obtained against the great majority of faults that can occur in the measuring channels; a fault whereby the meter reading lay between the two limits is very unlikely. The system thus possesses in large measure fail-safe characteristics: almost any conceivable defect in the protective system causes the safety rods to drop.

Thyratron circuit

In the circuits denoted *B* in fig. 1*b*, use is made of a tetrode thyratron, type PL 5727, which has two grids. The circuit arrangement is shown in *fig. 2a*. The thyratron anode load is fed by an alternating voltage of 50 c/s frequency. The range of input voltages V_i within which the thyratron ignites, extends from —0.20 V to —1.6 V, corresponding respectively to the lower and upper limits of the detected quantity Q. When V_i is between these limits, the thyratron passes a current pulse during each cycle. This current energizes the relay, keeping its contact closed and thereby energizing the D.C. electromagnets from which the safety rods are suspended. If Q reaches the upper limit ($V_i = -1.6$ V), the potential on the second grid of the thyratron is such as to just prevent the thyratron from igniting. The relay then cuts out, the magnet current is interrupted and the safety rods drop. If Q drops to the lower limit ($V_i = -0.20$ V), a transistor circuit ensures that a negative voltage is applied to the first grid sufficient to cut off the thyratron, with the same result. Further details of the operation of this circuit are given in the caption to fig. 2.

The input voltage V_i must be negative with respect to earth. For this reason the neutron-flux monitoring channel, described in article IIIB, contains an earthed tapping on the output voltage-divider (see fig. 2 of article IIIB); the voltage between the negative terminal and the tap serves as V_i.

Relays with *make*-contacts are used which are in the energized state during normal running; if a relay should cut out, for whatever reason, the electromagnet current is broken. This is important, for it provides the fail-safe feature: most faults likely to occur in the safety system, including supply failures, lead to the current through the relay coil being cut off, and hence to the shutdown of the reactor. Another important point in this connection is that the relays are D.C. relays, energized from an A.C. source via a rectifying element (the thyratron). Short-circuiting between anode and cathode or between anode and grid will thus open the relay.

Several of the above-mentioned features are also to be found in the safety system for the research reactor at the Technische Hogeschool, Delft (see the article cited under [1])). This system is in fact a further development of the circuit for the KEMA reactor, and was designed in 1955. The system used in the Delft reactor is fail-safe to an even higher degree; it is also considerably more reliable, owing to the use of a reduced number of components and, in particular, of independent detectors. The time required to stop the reactor is shortened by choosing the frequency of the alternating voltage in the anode circuit of the thyratron as 2000 c/s; with the subcritical KEMA reactor the time delay between detector signal and rod motion may permissibly be as long as 0.5 sec, for which a frequency of 50 c/s is sufficient.

The following are some of the features contributing to increased safety and reliability. The PL 5727 thyratron is a "Special Quality" valve [3]), and thus very reliable in itself. Two base pins are provided for the second grid, and both are used (fig. 2*a*), thus minimizing the risk of a broken grid connection (in which case the valve would pass current when it ought to be cut off). The relay coil is wound with thin wire, and therefore the risk of wire breakage due to corrosion cannot be excluded; this is countered by providing the coil with a second winding (the "anti-corrosion winding" w', fig. 2*a*),

[3]) Philips tech. Rev. **18**, 181, 1956/57.

which is connected at one end only and gives a degree of cathodic protection against corrosion of the main winding in the case of leakage currents.

Situations in which the safety system enters into operation

We shall now briefly review the situations in which the safety system enters into operation.

2) *Reactor period too short.* A short period means that the neutron level is rising too rapidly, and is therefore dangerous. As mentioned in article IIIA, the period is measured by differentiating the signal from the logarithmic neutron-flux meter. In the interest of reliability this is also done via two mutually independent channels. Each channel can operate the safety circuit.

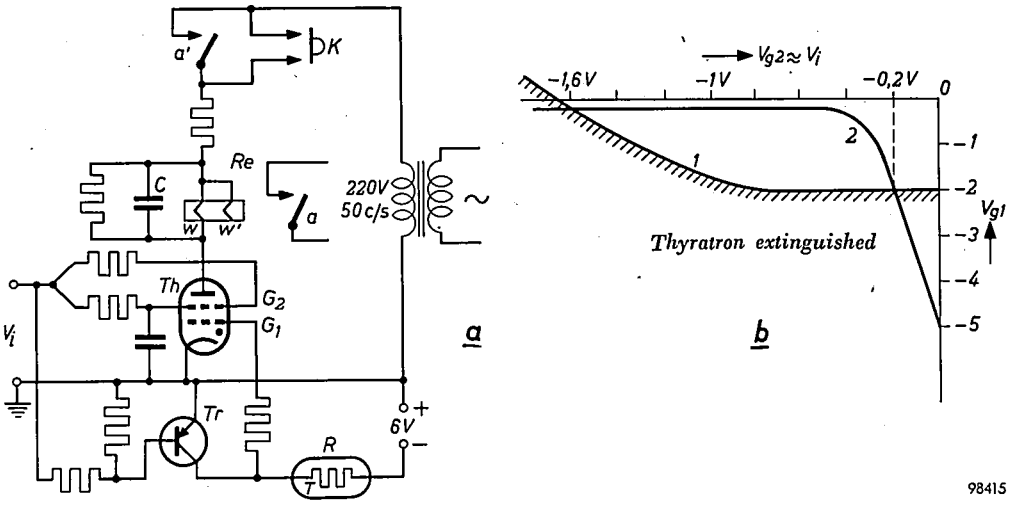


Fig. 2. *a*) Diagram of one of the identical safety circuits (*B* in fig. 1) together with relay *Re* (showing main winding *w*, anti-corrosion winding *w'*, contacts *a* and *a'*, and smoothing capacitor *C*). *Th* thyratron type PL 5727. *K* push-button re-set contact, in parallel with take-over contact *a'* of *Re*. *Tr* transistor type OC 71. The range of input voltages V_i in which the thyratron can ignite extends from −0.20V to −1.6 V, corresponding respectively to the lower and upper limits of the detected quantity *Q*.
b) Curve *1* gives the combinations of grid-voltage values V_{g1}-V_{g2} at which the thyratron, operated on a sinusoidal anode voltage of 220 V r.m.s., is just able to ignite. When the thyratron is extinguished, V_{g2} is virtually equal to V_i. Curve *2* shows the relation between V_{g1} and V_{g2} ($\approx V_i$) given by the transistor circuit.
Depending on the magnitude of V_i, three cases can be distinguished:
1) $V_i < -1.6$ V. In this case the transistor base voltage is sufficiently negative to allow the transistor to pass a high collector current I_c. The potential difference between emitter and collector is then very small, and therefore the potential on grid g_1 is practically equal to that on the cathode. Curve *1* shows that in this state ($V_{g1} \approx 0$, $V_{g2} < -1.6$ V) the thyratron passes no current.
2) $0 > V_i > -0.20$ V. The base voltage of *Tr* is now much less negative than in case 1). Consequently I_c is much lower and so therefore is the potential drop across the resistor *R*. Suppose that I_c is now zero: then the potential drop across *R* is also zero, and therefore $V_{g1} = -6$ V, whilst again $V_{g2} \approx V_i$. According to curve *1*, the thyratron again passes no current. However, I_c is not entirely zero but has a certain residual value, which increases steeply with the temperature of the transistor. V_{g1} is thus not so negative as −6 V, but if the temperature is not too high it is sufficiently negative to prevent the thyratron from igniting. To reduce the temperature effect, a resistor having a negative temperature coefficient is used for *R* (thermistor).
3) -0.20 V $> V_i > -1.6$ V. This is the case during normal safe running. V_{g1} and V_{g2} are now such that the thyratron can ignite when the anode circuit is closed by pushing the re-set button *K*. The relay is then actuated and closes the contacts *a* and *a'*, after which *K* can be released without the anode circuit being broken.

1) *Excessive neutron flux.* The neutron flux is measured via two entirely independent channels (see article IIIB), each with its own indicating instrument. As soon as the neutron flux reaches a level corresponding to 125% of the full-scale deflection, the upper-limit protection comes into operation. If the reading drops below 15% of the full-scale deflection, the lower-limit protection takes over.

3) *The voltage applied across the neutron detectors differs from the nominal value.* If, owing to some defect or other, the fixed potential on the neutron detectors should drop appreciably or fall off altogether, the signal from the neutron-flux channels will be too small to produce an adequate reading and therefore the lower-limit protection will come into operation.

Measures are also needed, however, to deal with the contingency of a relatively slight deviation of the voltage from its nominal value, since the voltage affects the sensitivity of the special neutron-flux detectors used (ionization chambers with some "gas amplification"). To avoid the use of highly sensitive and thus less robust voltmeters, the following method was adopted. The two neutron detectors are each fed from separate direct-voltage sources, the outputs from which are constantly compared. If the voltage difference exceeds 5 V, the safety system intervenes in the manner described below. The basic assumption is that the chance of the two voltages dropping simultaneously with a mutual difference of less than 5 V is negligible (unless both voltages fail, in which case, however, the lower-limit protection enters into operation, as described).

The difference between the two voltages (reduced by a small, variable zeroing voltage) is indicated by a voltmeter V (*fig. 3*), whose zero point is in the centre of the dial. This meter also functions as a relay [4]): when the needle reaches the one or other end of the scale, it intercepts a beam of light from an electric bulb, which is normally incident on a phototransistor. As soon as the latter receives no more light, the current it passes drops almost to zero, thereby actuating the safety system.

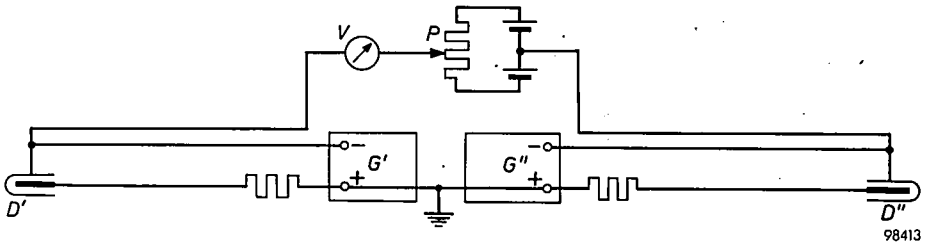


Fig. 3. Protective circuit that detects any deviation from the nominal voltage across the neutron detectors. It is assumed that the chance of both voltages dropping simultaneously at the same rate (i.e. differing by less than a few volts) is negligibly small. The one voltage thus serves as reference voltage for the other.
D', D'' neutron detectors, with voltage supplies G', G''. The difference between the voltages on D' and D'' (initially zeroed by a voltage derived from potentiometer P) is indicated by the voltmeter V, which also operates as a relay.

4) *The measured temperature differs from the preset value.* In article I it was explained that the reactivity of a suspension reactor has a negative temperature coefficient, which operates so promptly as to prevent a thermal run-away when the reactor is delivering power. The reactor under discussion, however, is used only for subcritical experiments (and hence has no significant power output). For the purposes of measurement it is provided with a temperature control system which automatically keeps the temperature in the reactor vessel constant within narrow limits (see article II of this series). It is conceivable that, owing to a defect in this control system or to improper handling, a drop in temperature might occur. Because of the negative temperature coefficient, this would be accompanied by an increase in the neutron flux. Although the above-mentioned measures exist to deal with excessive neutron level and a too rapid increase in neutron flux, for additional safety a further provision is made which brings the safety system into operation when the temperature differs more than 2 °C from the preset value.

5) *The suspension circulates too slowly.* If the flow rate of the suspension is too slow, the result may be an insufficiently homogeneous distribution of the fuel particles in the water. This gives rise to concentrations of fissile material that might prove dangerous. A flow-rate meter delivers a signal which, in the event of inadequate flow, actuates the safety system.

6) *Mains failure.* A mains failure deprives the thyratron anodes of their alternating-voltage supply. All safety relays then cut out and the magnet circuits are broken; the mains failure of course means in any case that the magnet current falls to zero.

Rod-position indication

Since the safety rods themselves are not visible, there should be some means of indicating whether they are still suspended from the electromagnets or have dropped into the reactor. In the KEMA reactor this indication is obtained in the following way. Use is made of the fact that the inductance of the magnet coils is higher when the magnetic circuit is closed by the safety rod than when it is open as a result of the rod being dropped. Each coil is bypassed by a capacitor (*fig. 4*) whose value is such that the circuit resonates at the ripple frequency (50 c/s) of their rectified supply voltage when the

[4]) See article mentioned under [1]), page 280.

rods are *not* suspended from the electromagnets; the ripple voltage across the coil is then considerable. A voltmeter connected to the circuit via small coupling capacitors gives full-scale deflection, in this case. When the rods are connected to the magnets, the ripple voltage is so much lower than at resonance that the meter gives only about 50% of full-scale deflection. We thus have the following cases:

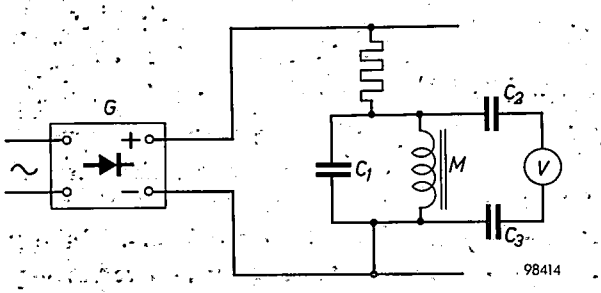


Fig. 4. Circuit for indicating the position of the safety rods. M electromagnet. C_1 parallel capacitor which causes the circuit M-C_1 to resonate at the ripple frequency of the output from rectifier G when the safety rod is *not* attached to the magnet M. V A.C. voltmeter. C_2, C_3 coupling capacitors.

Meter reading 50% . The safety rods are still attached to the magnets. The situation is normal.

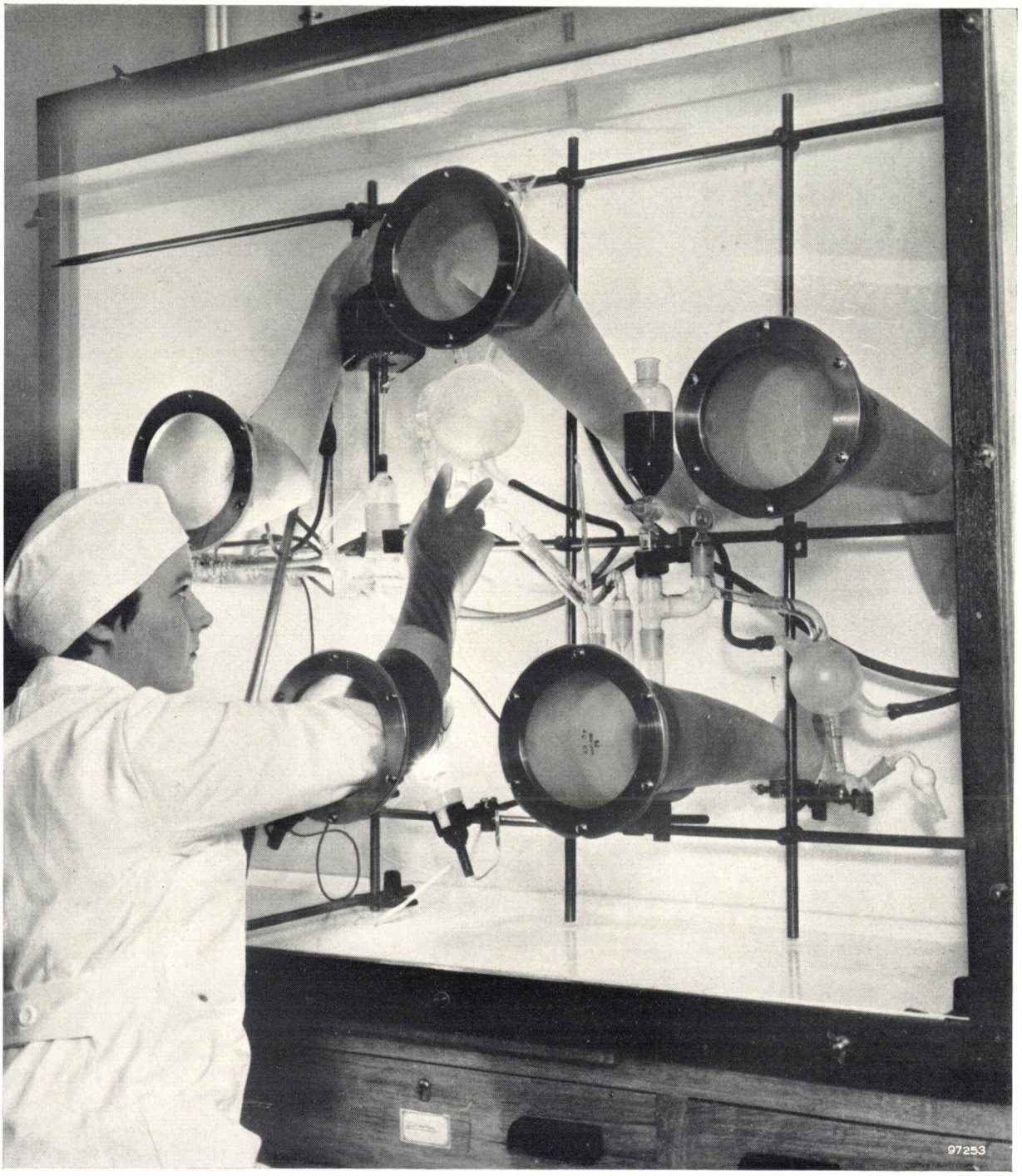
Meter reading 0 . . . No ripple voltage across the coil. The safety system has worked and the rods have dropped.

Meter reading 100% . The electromagnets are again energized, but the rods are not yet attached to them.

This meter, then, shows at a glance the position of the safety rods. If they have dropped, a pilot lamp (see fig. 4 of article IIIB) lights up to show which safety channel entered into operation first.

Summary. The KEMA subcritical suspension reactor contains three safety rods of boron carbide, which are suspended from electromagnets, the energizing current for which flows through a number of relay contacts in series. If one of the relays cuts out, the rods drop and the reactor is stopped. Each of the relays receives its energizing current via a thyratron, which can ignite only when its input voltage V_i is between −0.20 and −1.6 V, corresponding respectively to the lower and upper limits of the quantities to which the safety system is required to respond (neutron flux, reactor period, temperature, suspension flow-rate, etc.). The voltage V_i is taken from the channels used for monitoring these quantities. Measures are taken to make the system "fail-safe".

LABELLING OF COMPOUNDS WITH RADIOACTIVE TRACERS



The preparation of chemicals labelled with β-active isotopes must be done in such a way that the operator is not exposed to an atmosphere contaminated by dustborne or volatile radioactive substances). The work is therefore done in a "glove box", an air-tight box with a transparent window in which a number of long rubber gloves are sealed. The operator can thus manipulate the apparatus from outside. A small underpressure (5 cm water) is maintained in the box.

The photograph shows the operation of labelling tricresyl phosphate with the radio isotope ^{32}P, in the Isotope Laboratory of N.V. Philips-Duphar, Amsterdam. The preparation is used as a tracer in the oil industry.

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 Laboratories, Eindhoven, Netherlands.

2640: A. Kats: Spectres d'absorption du verre de silice et du quartz cristallin contenant des impuretés de Ge (Verres et Réfractaires **12**, 191-205, 1958, No. 4). (Absorption spectra of vitreous silica and crystalline quartz containing germanium; in French.)

The paper opens with an account of the investigations of Garino Canina relating to the absorption band at 2420 Å in vitreous silica, which have shown that this band is due to the presence of germanium. The author gives a detailed explanation of this phenomenon. The author's experiments on the irradiation of vitreous silica with X-rays have made it possible, with the aid of absorption spectra, to obtain precise data on the perturbations of the network and identification of the colour centres. The paper concludes with a table summarizing the results obtained on absorption bands and the types of colour centres formed by Ge and found not only in irradiated and non-irradiated vitreous silica but also in irradiated and non-irradiated crystalline quartz.

2641: J. A. Kok and M. M. G. Corbey: Aspects of electrical breakdown of liquid insulating material, I (Appl. sci. Res. **B 7**, 257-264, 1958, No. 4).

The electrical breakdown strength of insulating oil depends on the size of foreign particles which may form bridges in a place of maximum electric stress. The theoretical relation was verified with colloid suspensions of particles of known radius. Mineral oils may deteriorate if the particles unite by the process of flocculation, the occurrence of which depends on the relative magnitudes of the attractive London-Van der Waals forces acting between the particles and the repulsive forces between their ion atmospheres. The possibility of using a mineral oil as an insulator depends on the existence of an upper limit of the size of particle complexes due to the rapid fall-off of the L.-v. d. W. forces at diameters exceeding 500 Å. The latter diameter corresponds to a breakdown strength of 1 kV/mm. If acids are being formed, the upper limit of 500 Å will shift towards larger values, and correspondingly the breakdown strength may drop below 1 kV/mm.

2642: A. van Weel: Design of detector stages for signals with symmetrical or asymmetrical sidebands (J. Brit. Instn. Rad. Engrs. **18**, 525-538, 1958, No. 9).

The design of detector stages for signals with symmetrical-sideband components can be improved over conventional designs by properly using the long-established but not generally-known theory of such stages. For asymmetrical-sideband signals (e.g. television signals), an improved design is possible using the results of recent investigations. The condition for the i.f. amplitude curve to fall by a factor of two at the carrier frequency and for the v.f. section of the detector stage to have a wide-band transfer impedance are shown to be unjustified; it is shown that the correct design procedure is to consider the i.f. and v.f. sections of the detector stage together.

2643: J. Rodrigues de Miranda and J. J. Zaalberg van Zelst: New developments in output-transformerless amplifiers (J. Audio Engng. Soc. **6**, 244-250, 1958, No. 4).

An 11-watt audio power amplifier is described incorporating combined positive and negative feedback. The use of a Sinclair-Peterson type single-ended push-pull output stage enables an 800 ohm loudspeaker to be driven directly, providing excellent results. See also Philips tech. Rev. **19**, 41, 1957/58.

2644: P. Massini: Uptake and translocation of 3-amino- and 3-hydroxy-1,2,4-triazole in plants (Acta bot. neerl. **7**, 524-530, 1958, No. 3).

The uptake and translocation of radioactive 3-amino- and 3-hydroxy-1,2,4-triazole in various plants has been investigated. The first compound is taken up and translocated much faster than the second one upon administration to a leaf. The rate of uptake through the roots or through the cut-off stem is the same for the two compounds. The distribution of aminotriazole in the plants points to a translocation by the phloem system.

2645: W. J. Oosterkamp and J. Proper: The water equivalence of the phantom material Mix D for soft X rays (Brit. J. Radiol. **31**, 644, 1958; No. 371).

The phantom material "Mix D" (Jones and Raine,

Brit. J. Radiol. **22**, 549, 1949), Perspex, Philite and water have been compared as to their attenuation of X-rays generated between 23 and 50 kV constant potential. It is concluded that Philite (phenol-formaldehyde) approximates better than either Mix D or Perspex to water. (It has later been shown that the conclusion with regard to Mix D was not justified, the Mix D used not having had the prescribed composition; see W. J. Oosterkamp and J. Proper, Brit. J. Radiol. **32**, 560, 1959.)

2646: J. Brug and G. B. Paerels: Configuration of N-acetyl-neuraminic acid (Nature **182**, 1159-1160, 25 Oct. 1958).

The authors show that the condensation of oxaloacetic acid with N-acetyl-d-glucosamine as well as with N-acetyl-d-mannosamine in aqueous solution at *p*H 11 and at room temperature give rise to N-acetyl-neuraminic acid. It is shown that the N-acetylamino sugars epimerize under the conditions of the reaction.

From these results the authors corroborate the assumption of Comb and Roseman that the configuration of carbon atoms 5-8 in N-acetylneuraminic acid corresponds with the configuration of carbon atoms 2-5 in N-acetyl-d-mannosamine.

2647: A. Bril, H. A. Klasens and Th. J. Westerhof: On cathodo-thermoluminescence (Physica **24**, 821-827, 1958, No. 10).

The light output of some phosphors has been measured as a function of temperature from —180 up to 20 degrees centigrade. When the phosphors are measured in a demountable tube, strong maxima and minima can be found in the light output during heating of the phosphor from —180 °C to room temperature. This has also recently been described by Gobrecht, Hahn and Scheffler, who called this effect cathodo-thermoluminescence. A series of sealed and thoroughly outgassed tubes has been made. In these tubes, however, no indication of strong maxima and minima was found. The experiments lead to the conclusion that the cathodo-thermoluminescence does not seem to be a bulk property of the phosphors, but that it is more likely to be connected with gas adsorption on the phosphors.

2648: H. G. van Bueren: Plastic creep of germanium single crystals (Physica **24**, 831-837, 1958, No. 10).

The temperature and stress dependence of the creep rate of germanium single crystals, loaded in tension, have been measured. Three parts are distinguished in the creep curve. A theory is presented in which the initial part is explained in terms of the liberation of dislocation sources from impurity atoms, whereas the stationary creep rate attained in the second part is determined by the mobility of the dislocations themselves. This mobility is restricted owing to the occurrence of broken valence bonds along the dislocations. The third, work-hardening part of the creep curve is reached when the dislocation density in the material has reached a sufficient magnitude for piled-up groups to form.

2649: W. J. Oosterkamp: Meting en beperking van röntgenbestraling, ontvangen door onderzoekers bij doorlichting en het maken van foto's (Mens en Onderneming **12**, 387-390, 1958, No. 6). (Measurement and limitation of X-ray dose received by investigators during radiology; in Dutch.)

Review of the sources of radiation hazard for persons concerned with medical radiology and of the measures that may be taken to measure and limit the hazard.

2650: H. A. Klasens: The intensity dependence of photoconduction and luminescence of photoconductors in the stationary state (Phys. Chem. Solids **7**, 175-200, 1958, No. 2/3).

The properties of photoconductors and photoconducting phosphors are determined by states in the forbidden zone. Two models are considered for discussing the intensity dependence of their photoconduction and luminescence. In each model two recombination processes between electrons and holes are competing. It is shown that the one-state model, where direct recombination may take place between free electrons and free holes or via one impurity state, is generally not adequate. The two-state model with two discrete states having different capture cross-sections for capturing electrons and holes, can explain satisfactorily many aspects of these photoconductors quantitatively, such as superlinearity of luminescence and photoconduction, changes in power from 1 tot 0.5 in the conductivity versus intensity curves, activation and killing of phosphors, saturation, etc.

Duboc's method of considering all possible combinations between simplified versions of the equations describing the stationary state was applied successfully to give a mathematical description of the two-state model. Although there are 128 such combinations, all of which may occur under appropriate conditions, only a few occur in each individual case when only the intensity of excitation is varied. Others may appear when the temperature is varied. A few practical cases are analysed and the correct combinations of approximations determined. In particu-

lar, two examples of superlinearity are studied in detail. It appears that the general features of the curves can in both cases be described adequately in terms of the model with two discrete states, without having to introduce energy distributions of "ground states", as proposed by Rose. However, a problem remains in that the range of light intensity over which superlinearity occurs is much greater than is predicted by the simple model; it is possible that this can be explained e.g. by an inhomogeneous distribution of photoconductive excitation through the sample, due to strong absorption of the incident radiation.

2651: H. C. Hamaker: De recente ontwikkeling in proefopzetten met kwantitatieve factoren (Statistica neerl. **12**, 201-212, 1958, No. 4). (Recent developments in the design of experiments with quantitative factors; in Dutch.)

This paper reviews some recent developments in experimental designs for dealing with quantitative factors. The following topics are discussed: rotatable designs (Box), evolutionary operation (Box), experiments with mixtures (Scheffé), experiments with 10 to 100 factors (Satterthwaite), random vector experiments, the use of electronic computers in fitting non-linear models, and research strategy (McArthur).

2652: G. Meijer: Infrared fluorescence of copper-activated zinc sulphide phosphors (Phys. Chem. Solids **7**, 153-158, 1958, No. 2/3).

An infrared fluorescence was observed for ZnS:Cu,Al and ZnS:Cu,Ga at 1.57 μ and 1.72 μ. This emission appeared to be excited at low temperatures by radiation of 1.38 μ and of 0.5-0.8 μ, provided that the phosphor was excited by 3650 Å. The infrared quenching of the visible fluorescence and an infrared stimulation of the visible phosphorescence show a spectrum analogous to the infrared excitation spectrum, and seem to be connected with the same transitions.

2653: H. C. Hamaker: Some basic principles of sampling inspection by attributes (Appl. Statistics **7**, 149-159, 1958, No. 3).

Discussion of various approaches to one of the main problems in sampling inspection — that of deciding what size of sample to use. In particular, this paper discusses the extent to which our choice of sampling-inspection plans depends *a*) on knowledge of the distribution of the percentages defective in the inspection lots and *b*) on economic considerations. The opinion is expressed that these factors have a decided but crude influence, and that attempts to arrive at a precise economic theory of sampling-inspection procedures are doomed to fail in practice.

2654: M. J. Sparnaay: Corrections of the theory of the flat diffuse double layer (Rec. Trav. chim. Pays-Bas **77**, 872-888, 1958, No. 9/10).

The potential distribution in the diffuse part of the ionic double layer, which is formed at the interface between a solid and an electrolyte solution, is given by the Poisson-Boltzmann equation. This equation, involving some approximations, is modified such that now the following effects are included: 1) The effect of ionic sizes. 2) The effect of the dependence of the dielectric constant upon the ionic concentration and upon the electric field present near the interface. Two opposed tendencies can be distinguished here: *a*) the solution near the interface is polarized (Prigogine, Mazur and Defay), usually giving rise to a *de*crease of the ionic concentration near the interface; *b*) Coulomb interactions near the interface, between the ions and the interface, are different from those predicted from the unmodified theory. This will give rise to a relative *in*crease of the ionic concentration. It appears that tendency *b*) outweighs tendency *a*). A flat interface and a 1-1 electrolyte solution are considered only. The effects are represented as correction terms added to the original Poisson-Boltzmann equation. The new equation is integrated using elementary methods. The result of the calculations is that corrections are negligibly small when the electrolyte concentration is of the order of 10^{-3} normal. When the concentration is 10^{-2} normal, the corrections amount to 10-20% of the value of the double-layer potential, whereas the treatment breaks down, when the concentrations are 10^{-1} normal or higher.

VOLUME 21, 1959/60, No. 6 pp. 157-184 Published 15th March 1960

Philips Technical Review

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

THE USE OF RADIOACTIVE ISOTOPES FOR THE STUDY OF LITTORAL DRIFT

by J. J. ARLMAN *), J. N. SVAŠEK **) and B. VERKERK ***).

621.039.85:627.223.7:627.521.1

In coastal and fluvial engineering works it is sometimes necessary to know in what way and in what quantities sand and mud at a given location will be transported by currents and waves. This problem is of great significance in the Netherlands Delta Project. The comprehensive programme of measurements set up by the Dutch Rijkswaterstaat **) *to follow the gradually changing hydrographic situation around the coast includes direct measurements of sand movements by means of radioactive tracers. Compared with other tracer applications this involves the use of large amounts of radioactive materials; careful preliminary investigations were therefore required to decide on questions of procedure and radiation protection.*

Introduction

The movement of sand and mud by water currents has been a factor of vital importance to man's settlements on the coasts since the earliest times. As a result of silt accretion from rivers and ocean current, harbours like Pisa and Bruges have come to lie far inland, whilst others, whose names are now scarcely remembered, such as Reimerswaal, have been undermined by erosive currents and have vanished into the sea. Others, too, have been preserved from the consequences of coastal detrition and siltation by human intervention. An illuminating example is Venice, of which the story goes that, by diverting the course of the River Brenta, which threatened to silt-up the lagoon, it was able to preserve its strategic position as an island city and port — at the expense of its vassal township Chioggia. However, we do not need to look so far back for examples of works for controlling sand and mud movements. The position of the Netherlands largely depends on such works, whether they concern land reclamation, as in the north of the country, or, as in the south-west, the protection of the coast and the keeping open of waterways by piers, groynes, etc.

It may well be that the execution of the Delta Project will call for new and extensive works of this kind. The envisaged damming of the estuaries (*fig. 1*), already begun, will reduce the length of the dykes to be defended against the tides from 1700 to about 1350 kilometres. At the same time, however, the damming will cause drastic changes in the hydrographic situation of this coastal area, particularly in the "underwater delta" adjoining the mouths of the estuaries. The currents in and out of the mouths, at right-angles to the coastline, will largely disappear, but the remaining currents parallel with the shore will bring about sediment displacements different from those known hitherto. The resultant change in the depth contours will in turn alter the configuration of the currents and also the wave phenomena, so important to sand and shingle movements, on the coast. The altered situation must at all costs be prevented from leading to the undermining and disintegration of the sand-dune foreshores, or blocking of the inlets to the Westerschelde and Rotterdam waterways.

The expected changes in sand movements along

*)N.V. Philips-Duphar Isotope Laboratory, Amsterdam.
**)Rijkswaterstaat, Delta Authority, The Hague. The Rijkswaterstaat is a national body responsible for canals, drainage, water conservancy and coastal protection.
***)Now at the Netherlands Reactor Centre, formerly of Philips Research Laboratories.

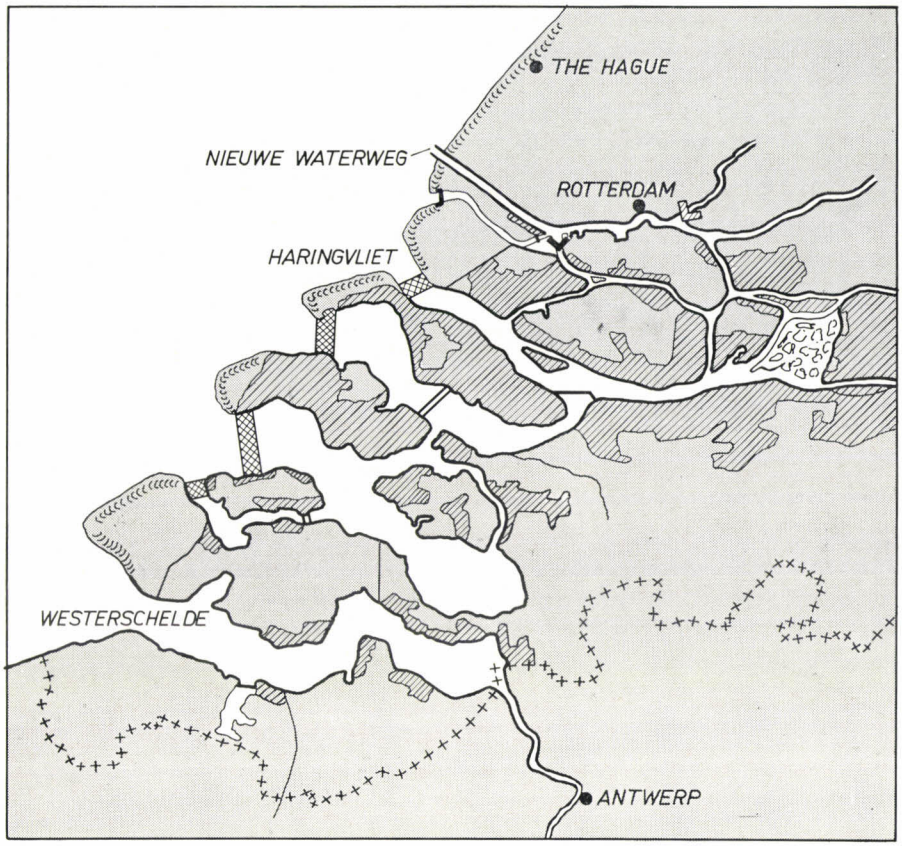


Fig. 1. Survey of the Dutch Delta Project, showing the projected damming of the estuaries. The damming of the Haringvliet is already in an advanced stage. Dykes that are at present still exposed to the action of tides are outlined in bold lines. The hatched parts indicate the areas inundated during the flood disaster in 1953.

Sand dunes along the coast.

Projected dams closing the estuaries.

Storm gate in the Hollandse Yssel.

Dams needed for implementing the coastal protection works.

Dams designed to improve the estuary conditions.

Storm gate in the Oude Maas, open during storm tides.

the south-west coast can be predicted to some extent on theoretical grounds. The picture presented by such predictions is not unfavourable, but the uncertainties involved are very considerable. The transport of sediment, which takes place partly on the river and sea beds and partly in the water itself (where sediment particles are present in suspension) is a much too complex phenomenon for exact calculation, being governed not only by permanent and quasi-permanent two-dimensional currents and by wave action but also by local disturbances, turbulence, the Coriolis force and, at bends, by centrifugal forces. Again, the configuration of currents and waves themselves, being determined by numerous factors including the tides, fluvial discharges, differences of density in the water and winds, is something that can only be learned from measurements [1]).

For the purposes of the Delta Project it was therefore necessary to make an extensive and largely experimental study of littoral drift in the coastal areas concerned. The investigations are in part being carried out on scale models in hydrological laboratories. A large model of the Haringvliet (horizontal scale 1 : 150), which will be the first of the sea inlets to be closed, has been the subject of experiments for some considerable time now at De Voorst, and another model, covering a coastal belt from the Haringvliet to beyond the mouth of the

[1]) For a further discussion of this question see the Delta Authority's Progress Report of June 1957 (also published as Technical Memorandum No. 105 of the Beach Erosion Board, March 1958): J. J. Arlman, P. Santema and J. N. Svašek, "Movement of bottom sediment in coastal waters by currents and waves; measurements with the aid of radioactive tracers in The Netherlands." Many of the considerations and experiments described here are discussed at greater length in this report.

Rotterdam Waterway (horizontal scale 1 : 800), was completed and taken into use in 1958 at Delft. Moreover, since various factors, such as wave action, cannot be simulated exactly in such models, "field experiments" are being carried out during the Delta works and will be continued after their completion. These experiments fall into three categories.

First, the *causes* of the sand movements are measured, namely currents and waves (which must also be known for other reasons). The measurements are made at fixed times in numerous places simultaneously, and constitute snapshots, as it were, of the current and wave conditions (height and spectral distribution of the waves) prevailing at given instants in the entire region under investigation. Extensive use is made of modern methods, such as the automatic transmission by radio of recorded data from fixed measuring points to an information centre (telemetering) and charting the position of test boats, of freely drifting rafts and even of wave crests by means of an 8 mm radar installation [2]).

The second category of measurements concerns the *consequences* of the sand movements: in a number of 500 metre strips along the coast extending from the foreshore to a depth contour 10 m below the mean sea level (Amsterdam datum level) (*fig. 2*), changes of the coast profile are studied from monthly soundings and analysed samples of bottom sediment.

In the third category the *actual process* of the sand movement is measured directly. This was made a practical possibility by a highly effective measuring technique developed during the last decade or so, and based on the use of radioactive tracers. When the sand at a particular site of the river or sea bed is "labelled" with such a tracer it is possible, owing to the high sensitivity inherent to radioactive measuring techniques, to detect some time later the presence of radioactivity in the sand at fairly considerable distances from that site, and in this way to track the movement of the sand. Quantitative data on sand transport can be obtained in this way. Details of this method and the problems it involves, some of which have been jointly investigated by the Delta Authority under the Dutch Rijkswaterstaat and the Isotope Laboratory of N.V. Philips-Duphar, will be discussed in the present article.

Other and essentially simpler methods exist for directly measuring littoral drift. There is the "Sfinx" meter for the movement of bottom sediment and the "Delft flask" for the transport of particles in suspension, both based on the collection of moving sand in a calibrated vessel [1]). The results obtained with these and similar contrivances show very wide spreads, however, partly because the setting-up of the instrument affects the local state of flow in a manner that is difficult to ascertain. A further drawback is their limited usefulness in rough weather, precisely when sand movements may be most pronounced.

Principle of the method

A physical or chemical process in a given material can be investigated by means of a "tracer" of any kind, provided the tracer behaves in regard to the process in the same way as the material in question. In our case, then, the tracer must be a granular material that is transported by water in the same way as sand, or, to be more exact, as the sand found in the coastal area of the Delta works.

Supposing we already have such a tracer, the method of going about the measurements will be roughly as follows. A quantity of the tracer, possibly mixed beforehand with the ordinary unmarked sand, is carefully dumped at a given moment on to a certain site of the sea bed. Currents and wave action will now spread out this "tracer bed" over a large area. The requirement is that the tracer should later be detectable up to a distance of, say, 1000 metres from the dumping site. For the sake of argument we assume that the tracer has spread uniformly over a circle of 1000 m radius, thereby mixing with unmarked bed material to a certain depth d of, say, 10 cm. In that case the tracer concentration has been diluted by an amount of 300 000 cubic metres of unmarked sand.

Plainly, then, the measurements must be extremely sensitive in order to detect the tracer, and in our case, therefore, only radioactive tracers enter into consideration [3]). Even then, using the most sensitive of detectors, it remains necessary to work with enormously high levels of radioactivity. The following rough calculation will make this clear.

---

[2]) A similar 8 mm system is described in: J. M. G. Seppen and J. Verstraten, An 8 mm high-resolution radar installation, Philips tech. Rev. 21, 92-103, 1959/60 (No. 3). A provisional communication on the use of radar in this connection will be found in the quarterly report "Deltawerken" No. 4, May 1958, pp. 11-19 (in Dutch).

[3]) In principle the tracer used can also be a substance that differs from sand in its chemical or mineralogical properties, in its colour, or in its property of fluorescence. Quite apart from the inadequate sensitivity of these methods of analysis when the tracer is diluted to the extent mentioned above, or the expense involved, a drawback of such tracers is that samples from the sea or river bed have to be taken (which is *not* necessary with radioactive tracers). This makes measurements at numerous points a laborious and time-consuming process. Furthermore, repeated measurements in the same area are not possible with such tracers, since the tracer properties of the material remaining behind in the sediment (although very diluted) do not decay as is the case with radioactive tracers. See the account of the measurements below (pp. 160-166).

Because of the admixture with inactive material up to a fairly considerable depth (d), the tracer material used must be a radioactive isotope that emits radiation of great penetrating power, e.g. gamma rays of 1 MeV or higher. The best detector for such hard radiation is the scintillation counter, using, for example, an NaI crystal [4]). This counts a substantial fraction of the hard gamma quanta incident on the crystal (the counting efficiency η may be 35%), and discriminator circuits can be used to reduce the nuisance of background radiation. The measurement is carried out with a set-up such as illustrated in *fig. 3*. The counter, containing a crystal of e.g. 2.5 cm diameter, is positioned at a

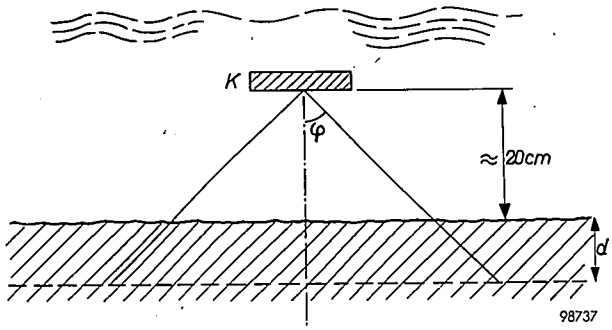


Fig. 3. Detection of bottom-sediment radioactivity by means of a submerged scintillation counter (NaI crystal K). It is assumed that the tracer is uniformly mixed with the inactive sea sand over a depth d.

distance of, say, 20 cm above the sea bed. In every cubic centimetre of bed material there are n radioactive grains, each with an activity of a microcuries (1 microcurie corresponds to the activity of 10^{-6} grams of radium, i.e. 3.7×10^4 disintegrations per second), and the isotope emits p separately detectable gamma quanta per disintegration. The detector then has a count rate of

$$C = 3.7\times10^4\times60\ pna\eta G \text{ pulses per minute.} \quad (1)$$

The factor G is a "geometry factor" which takes into account the dimensions referred to above, the semi-angle φ subtended at the apex of the cone of radiation reaching the counter and which we take to be, say, 45°, the depth d in the sea bed over which the radioactive grains are distributed, and finally the absorption suffered by the radiation on its way to the counter. For a given radiation and at $d = 9$ cm, the count rate has been calculated in this way [1]) as

$$C = 7\times10^6\ na \text{ pulses per minute.}$$

With the scintillation counter a background of about 300 pulses per minute is measured and for radioactivity to be detected with certainty it must give a count rate of at least 50% of the background, i.e. in the present case 150 pulses per minute. This means we must make $na \geqq 2\times10^{-5}$ microcurie/cm^3; if the tracer is diluted by 300 000 m^3, as assumed above, the total activity required at the dumping site is about 6 curies (corresponding to 6 grams of radium). If the tracer is mixed to a greater depth, say $d = 1$ metre, the figure is much higher, in the present case about 40 curies.

There are many other requirements to be met by the tracer material, important ones being the half-life of the isotope, the activity per grain and the nature of the grains. We shall return to these presently. The fact, however, that the measurements would involve such very high levels of radioactivity made it desirable to investigate at an early stage the methods to be adopted in performing the measurements and in handling the tracer material, and above all to consider carefully the means of effectively safeguarding the personnel against harmful radiation. It was therefore decided that Delta Authority engineers, in cooperation with Philips-Duphar, should first carry out a preliminary experiment in the De Voorst hydrological laboratory, which is situated on sandy ground unsuitable for cultivation in the North-East Polder near Vollenhove. A brief description follows of the experimental arrangements and procedures.

Preliminary estimates were of course made of the radioactive contamination likely to be caused by the coastal measurements in nearby bathing beaches and fishing waters. The dose rates to be expected in these places were calculated from the above-mentioned degree of dilution and from data on the half-life and activity per grain of tracer material (see end of article). On the strength of these calculations, and having regard to the safety requirements — which were made about 1000 times more stringent than in former experiments of this kind elsewhere [5]) — the proposal to perform the measurements along the coast was approved by the Inspectorate of Public Health and by the Isotope Committee of the Royal Netherlands Academy of Science.

Preliminary experiment in the North-East Polder

The preliminary experiment was made in an "artificial river" specially designed for studying sand and silt movements. The river consists of a straight channel 39 m long and 2.5 m wide, with brickwork banks and sandy bed and with adjustable weirs at both ends for varying the water level

[4]) See e.g. J. A. W. van der Does de Bye, The scintillation counter, Philips tech. Rev. **20**, 209-219, 1958/59 (No. 8).

[5]) J. L. Putman, D. B. Smith, R. M. Welles, F. Allen and G. Rowan, Thames siltation investigation, preliminary experiment on the use of radioactive tracers for indicating mud movements, Joint A.E.R.E.-D.S.I.R. report, Wallingford, Dec. 1954.

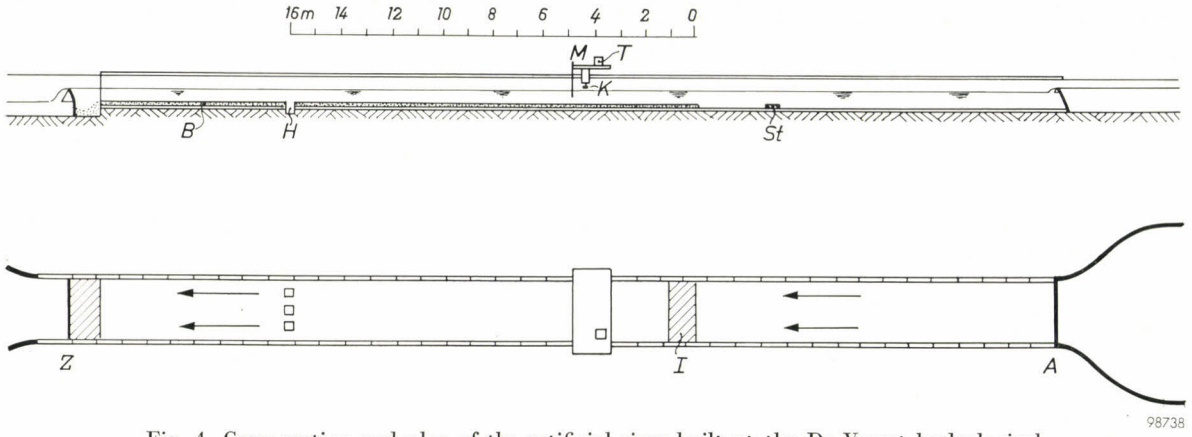


Fig. 4. Cross-section and plan of the artificial river built at the De Voorst hydrological laboratory in the North-East Polder. *A* water inlet, *Z* water outlet, *B* sand bed, *I* site where tracer material was dumped, *M* sliding bridge with scintillation counter *K* and instruments *T*, *H* sand traps for independent determination of sediment movement. At *St* a standard sample of tracer material was fixed and measured simultaneously during each test run, in order to allow for the gradual decline in tracer activity.

and the rate of flow (*fig. 4*). A scintillation counter, surrounded by a 300 kg lead shield to reduce interference from undesired radiation, was suspended from a wooden bridge capable of being slid along the banks. On top of the bridge were mounted the pulse-counting equipment (*fig. 5*), a dip-stick for reading the water level above the bed and a tube for drawing up samples of bottom sediment.

When the experiment was planned, similar experiments had already been carried out at two other places. The first concerned an investigation of

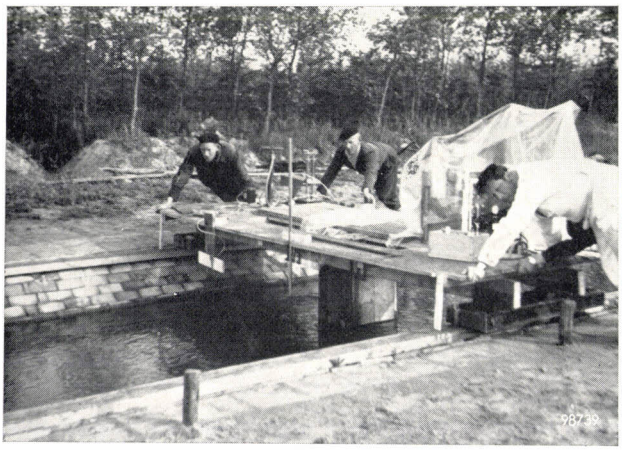


Fig. 5. Bridge with detector and instruments, used for the preliminary experiment at the De Voorst hydrological laboratory.

siltation in the Thames, and was made by members of the Atomic Energy Research Establishment, Harwell, in collaboration with the Port of London Authority [5]). Almost concurrently, Japanese investigators studied littoral drift along Hokkaido Island with a view to the building of a large harbour near the town of Tomakomai [6]). Since there was no possibility of suitably making sand grains themselves radioactive, use was made in both instances of powdered glass in which a radioactive isotope was fused (in England scandium, ^{46}Sc, and in Japan zinc, ^{65}Zn). Although it was to be predicted that the transport characteristics of glass grains would differ considerably from those of sand grains, a similar tracer was nevertheless chosen for our preliminary experiment, namely glass beads. In this case, however, instead of fusing an artificially radioactive substance into the glass, the glass beads were themselves made radioactive by irradiation in a nuclear reactor. Under neutron bombardment the radioactive isotope ^{24}Na is produced from the sodium in the glass; this isotope emits very hard gamma rays (1.38 and 2.76 MeV), and has a half-life of about 15 hours. This half-life is much too short for the large-scale investigations in the Delta area, where some measurements will extend over several weeks. For the preliminary experiment, however, a short half-life was just what was wanted in order to avoid the danger of prolonged radioactive contamination of the laboratory premises.

The tracer material was obtained by irradiating fifteen small aluminium cans, each containing 128 grams of glass beads, in the Dutch-Norwegian nuclear reactor at Kjeller, near Oslo. The irradiation lasted about a week, at the end of which time a total activity of 20 curies was measured (partly from shorter-lived isotopes than ^{24}Na).

Conveying the material from Kjeller to De Voorst was a problem in itself; it had, of course, to be done quickly, in view of the short half-life of ^{24}Na, and

[6]) S. Inose, M. Kato, S. Sato and N. Shiraishi, The field experiment of littoral drift using radioactive glass sand, Proc. Conf. Peaceful Uses Atomic Energy, Geneva 1955, Part **15**, p. 211.

at the same time thorough measures of radiation protection were necessary. Air transport was ruled out as too expensive. Road transport was therefore decided on. The containers were enclosed in a lead shield weighing 700 kg, and the truck was manned by three drivers who drove in shifts all through the day and night. The customs posts at the borders to be crossed (Sweden, Denmark, Germany) had been informed of the plan and gave every assistance to prevent delays. As a result, in spite of a breakdown on the road, there were still 1.5 curies of ^{24}Na left upon arrival at De Voorst.

Elaborate precautions attended the unloading of the trucks. A person standing about three feet away from the unshielded cans would have received in six minutes the maximum permissible dose for a whole week; moreover the special facilities available for handling radioactive material at the place of departure were lacking at De Voorst. The material was unloaded in an improvised dump consisting of a trench 50 metres long, access to which was barred on three sides by dense shrubs and brushwood. The cans were deposited at marked places a few yards apart so that in subsequent handling there was only the radiation from one can to be reckoned with. The contents of the first can were mixed with a known quantity of sand and water in a concrete-mixer, which was mounted on a plastic sheet for collecting spilt material. With a few simple tools the can could be opened in the mixer by remote operation. The labelled sand was then immediately conveyed to the artificial river and emptied out at the appropriate site (see *fig. 6*).

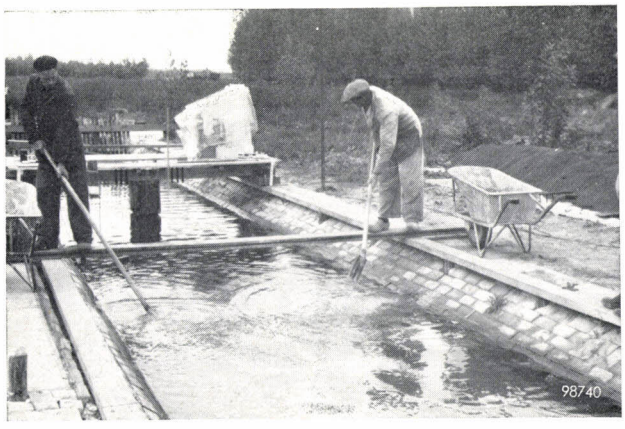


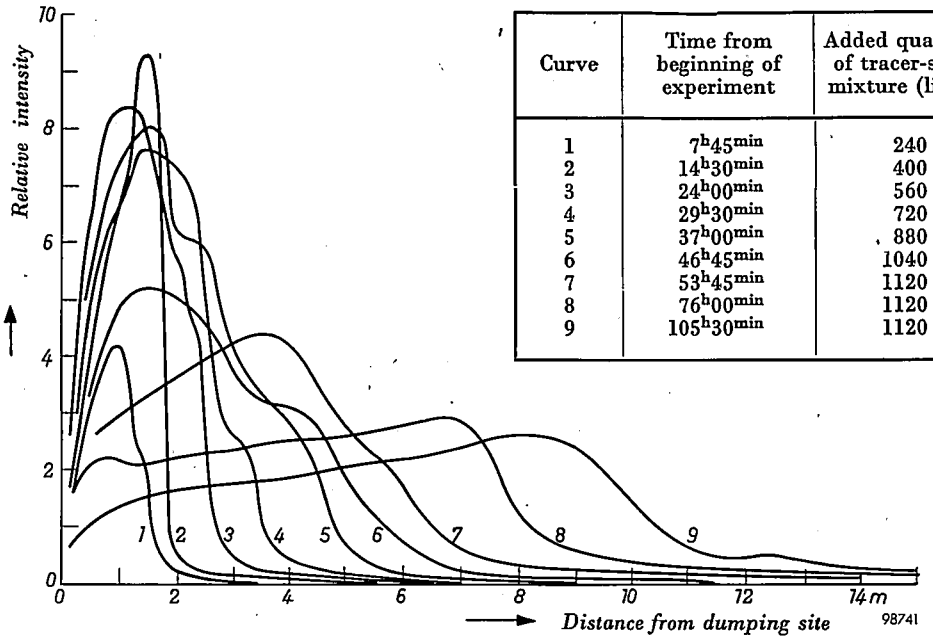
Fig. 6. Dumping the tracer in the artificial river.

This having been done, measurements were made at regular intervals of the distribution of activity along the axis of the channel over a length of about 15 metres. All results were reduced in the usual manner to percentages of a standard sample of the original radioactive material which was measured simultaneously, in order thus to correct for the gradual decline in radioactivity with time. A few hours later the same quantity of sand, labelled with glass beads from the second can, was dumped at the same site, and the activity measurements were then resumed. After 50 hours all the radioactive material had been used up in this way. The measurements were carried out for a total of 100 hours from the first dumping.

The results of the activity measurements made at various times over the length of the channel are represented in *fig. 7*. The broken curves in *fig. 8* show the rates of transport derived from these results (in litres of bottom sediment per hour), compared with the true rates of transport (full curves) which, because of the simple geometry of the channel, could be calculated from the measured changes in the configuration of the bed. It can be seen from the latter curves that at the beginning of the experiment there was virtually a steady-state movement of sediment; the quantities of sand carried away at the end of the length of channel tested were roughly equal to the quantities supplied to the dumping site. At this early stage, however, the transport of the glass beads clearly lagged behind that of the sand. As mentioned, a difference in this respect was to be expected: the average size of the glass beads was greater than that of the sand grains (D_{50}, that is the value of the diameter which is exceeded by 50% by weight of the material, was 270 μ for the glass beads as against 160 μ for the sand at De Voorst). Moreover, the size distribution of the grains was entirely different, and the specific gravity of the glass was 2.95 as against 2.65 for the sand. As can been seen from fig. 8, after the first 50 hours, when there was no longer a steady-state movement of sediment, but bed erosion or "scour" had set in, transport of the glass beads occurred. Previous to this, they had evidently largely remained behind with the larger of the sand grains.

Apart from giving a clear idea of the errors to be expected if the tracer material does not possess the same transport characteristics as the sand, the experiment demonstrated that the measuring technique employed can yield useful information on the distribution of radioactivity present in a given area, even when the activity is low (after 100 hours there remained only 1.5% of the original activity).

Numerous measures were taken to protect the personnel against radiation during the experiment. The radiation hazards involved in the various operations were checked by systematic monitoring. In



Curve	Time from beginning of experiment	Added quantity of tracer-sand mixture (litres)
1	7^h45^{min}	240
2	14^h30^{min}	400
3	24^h00^{min}	560
4	29^h30^{min}	720
5	37^h00^{min}	880
6	46^h45^{min}	1040
7	53^h45^{min}	1120
8	76^h00^{min}	1120
9	105^h30^{min}	1120

Fig. 7. Tracer activity as a function of distance from dumping site, measured at various times along the axis of the channel (fresh tracer material being added at regular intervals; see inset table).

some cases, for example during the operation of the concrete-mixer and when dumping the active sand, continuous monitoring was adopted. All persons engaged in the tests, including some unskilled workers, were of course thoroughly instructed on procedure beforehand. All wore film badges for checking the total received dose, and those most exposed also carried pocket dosemeters [7]) for measuring dose rates at appropriate moments. The area in which radioactive material was handled was plainly marked off by warning signs, and persons unconnected with the experiment were not allowed to enter. Radioactive contamination of the ground near the concrete-mixer could not entirely be avoided, but since the personnel wore either clogs or gum-boots which were left behind at the end of the day, no radioactivity was trodden outside. Moreover, the short life of ^{24}Na meant that all contamination of the premises had disappeared after about a week.

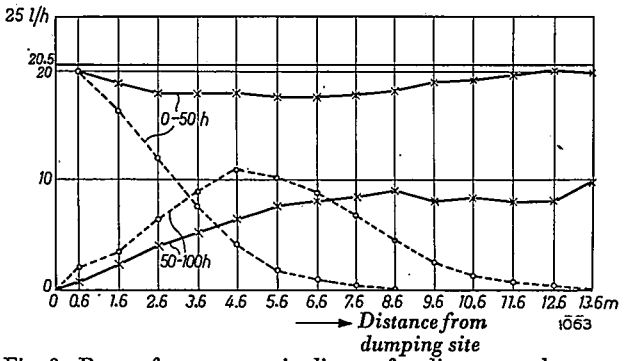


Fig. 8. Rate of transport, in litres of sediment per hour, at points along the axis of the channel in the first 50 hours of the experiment (during which fresh tracer material was repeatedly added at an average rate of 20.5 l/h) and in the following 50 hours (during which no further tracer was added). The broken curves are derived from the activity measurements, the solid curves from the changes in the configuration of the bed. The transport characteristics of the tracer (activated glass beads) differ considerably from those of the sand. This is particularly evident from the curves for 50-100 hours, which, in the first 6 metres, show a larger rate of transport of radioactive material than corresponds to the actual rate of sand transport there: initially, in this first 6 metres or so, the tracer grains were hoarded up, as it were, and only released when there was not much normal sand left over.

At the end of the experiments it was found that only a few persons had received a total dose in excess of 100 milliroentgens. The highest dose was 160 mr, received by those who had worked 3 days for 8 hours a day. At that time the permissible weekly dose was 300 mr [8]).

Labelling the sand for measurements on the coast

As we have seen, the simple method of using activated glass beads is not suitable for the definitive measurements. The half-life of ^{24}Na is also far too short. What is wanted is a half-life of the order of one week to a month; there is then sufficient time to carry out measurements extending over a few weeks, whilst on the other hand the activity decays quickly enough to avoid any cumulative conta-

[7]) See e.g. N. Warmoltz and P. P. M. Schampers, A pocket dosemeter, with built-in charger, for X-radiation and gamma radiation, Philips tech. Rev. **16**, 134-139, 1954/55.

[8]) The weekly permissible dose has since been reduced to an average of 100 mr, but peaks are allowed provided that the total dose received in 13 weeks does not exceed 3 r.

mination of beaches or interference with subsequent measurements in the same area.

Direct activation of a quantity of sea sand, by irradiation in a nuclear reactor, would be the ideal way of ensuring completely identical behaviour of the tracer in regard to its transport by currents and waves. In particular there would then also be a reasonable certainty that each grain had received an activity proportional to its weight; in that case the measured activities would be a direct measure of the displaced quantities of sand by weight, irrespective of the segregation between grains of different size that may occur during displacement. Unfortunately, for our purposes the elements present in sea sand either produce isotopes of too low an activity or they lead merely to beta emitters, whose radiation is not penetrating enough, and which moreover are short-lived. It was therefore necessary to look for a suitable radioactive isotope, and to find a means of bonding it or incorporating it chemically in sand, or in a substance sufficiently like sand for our purpose.

An investigation into the possibly suitable isotopes led to the choice of ^{46}Sc. This has a half-life of 85 days and emits two gamma quanta, one of 0.89 MeV and one of 1.12 MeV. The presence of two radiations of different energy can be an important advantage, for in principle it is then possible, by virtue of the different absorption of these radiations in the bottom sediment, to deduce from the measurements information on the activity distribution in a vertical cross-section of the sea bed. (This implies, however, that the semi-angle φ in fig. 3 would have to be rather small, which would reduce the sensitivity of the activity measurements.)

The isotope ^{45}Sc, from which the active ^{46}Sc is formed by neutron-capture, has an effective cross-section for this process of 22 barns; this is a relatively high value, and means that when a quantity of the material is irradiated in a nuclear reactor a quite considerable activity per unit weight can be produced. In view of the high total activity needed in our case (as mentioned above, as much as 40 curies may be required) this is a favourable circumstance, for there is only limited space in a reactor for the parent material and strongly radioactive materials become more difficult to handle as their volume increases.

There is no known method of bonding ^{46}Sc to grains of sand [9]). An effective and economical alternative, however, is to bond the isotope to an inorganic ion-exchange substance. In connection with their normal function — the binding of ions from a solution which is passed through a column packed with the ion-exchange material — the ion-exchange substances are prepared in the form of grains. Some have grains of nearly the same size as the sea sand with which we are concerned. Furthermore, the exchangers are capable of binding ions of trivalent scandium so strongly that subsequent exchanges with monovalent or divalent ions in sea water are negligible. For that reason substances of this kind, particularly of the clayish zeolite group, are in fact used for "cleaning-up" the radioactive waste from nuclear reactors, which is deposited at inaccessible places. One of the zeolites known as "greensand", which is commercially available as "Ionac C 50", has a density (2.72-2.76) close to that of North-Sea sand (2.65-2.68) and it can readily take up the necessary quantity of scandium.

The take-up capacity of Ionac C 50 cannot by any means be used to the full. Indeed, the activity per unit weight of tracer, that is to say the activity per grain of zeolite, must not be unduly high. There are two reasons for this. Firstly, they must not be injurious to humans or fish that may swallow some. Secondly, the activity must be carried by so many grains that even in the most diluted state of the tracer the statistical fluctuations in the number of grains present in the "field of view" of the scintillation counter will remain sufficiently small. Referring back to the foregoing calculation (see formula (1) and discussion following it), we can put the required number of grains for one measurement at 100. At the beginning of the experiment described we should then need a total of approximately 2.5×10^9 grains. If the average grain diameter is 200 μ, this means that the total initial activity of about 6 curies must be contained in at least 25 kg of the ion-exchange substance.

The activity per grain is then low enough not to cause any harm to fish, even though they may swallow large numbers of the widely dispersed grains. Sea bathers would have to consume kilograms of sea sand before their activity intake from the tracer in its ultimate dilution reached the permissible level. *A fortiori* the activity to which bathers lying on the beaches might be exposed is far below the permissible limit.

One difficulty, which typifies the whole experiment, was that the grains of Ionac C 50 are *softer* than the sand grains in which they are mixed. During the movement of sediment over the sea bed the grains are therefore *eroded* by the sand, their

[9]) This has, however, proved possible with ^{140}Ba, an isotope that might also be used for this purpose inasmuch as its daughter product ^{140}La emits hard gamma quanta. See D. B. Smith and J. D. Eakins, Unesco Conf. Radio-isotopes in sci. Res., Paris, Sept. 1957, paper No. 63.

size distribution changes rapidly, and the resultant segregation again causes the transport properties of the tracer to differ excessively from those of the sand. It is possible, however, to make the Ionac C 50 grains harder by firing the material at 750 °C. The capacity of the fired material to absorb ^{46}Sc, and the rate at which it does so, are still sufficient, allowing the active isotope to be added to the Ionac C 50 *after* firing, which is of course the easier procedure; the limited activity required per grain is absorbed in about 2 hours (the unfired grains absorb the same activity in 3 minutes). The fact that firing adequately reduces the erosion is demonstrated in *fig. 9*: a mixture of fired Ionac C 50 and quartz

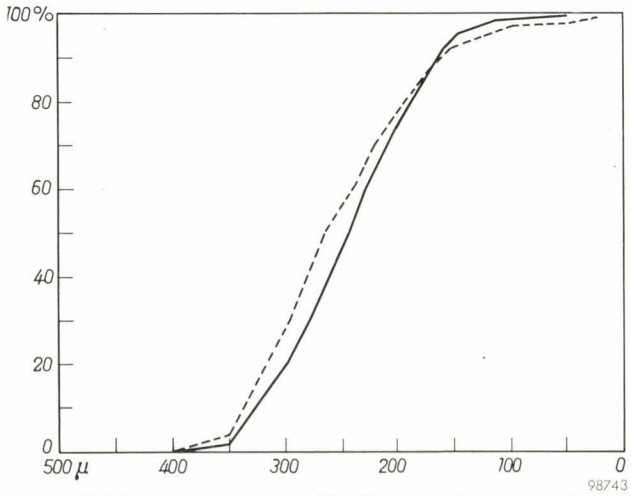


Fig. 9. Solid curve: grain size distribution of a given mixture of sand and fired "Ionac C 50" (the ion-exchange substance in which the radioactive isotope ^{46}Sc is absorbed), plotted as the percentage by weight of grains *larger* than the value on the abccissa. Broken curve: the distribution after the mixture, suspended in water, had been shaken up for 10 hours in a concrete-mixer. There is scarcely any change in the distribution, indicating that there is virtually no erosion of the fired Ionac C 50 grains by the sand.

sand was shaken up for 10 hours in a concrete-mixer, and the results show that the size distribution of the mixture remained reasonably constant. After the Ionac C 50 had been stirred in sea water for 48 hours, it was not possible to detect any drop in ^{46}Sc activity.

The measurements in the Delta coastal area are therefore being carried out with this tracer material.

The coastal measurements now in progress

In April 1958, trial measurements were made in the mouth of the Haringvliet (the tracer being dumped at point *A* in fig. 2). These provided an opportunity to test the method in the field, and at the same time yielded the first data on sand movements in this area, which is of particular importance to the Delta Project at the present stage.

Before the tracer material was dumped the area was scanned for radioactivity in order to determine the background to be allowed for during the measurements. This brought to light a number of unsuspected facts. The sand on the sea bed was found to possess much higher natural radioactivity than the sea water. This activity appears to be concentrated in the heavy minerals, which constitute from 1 to 5% of the sand, and is thought to originate from uranium or thorium inclusions in the mineral zirconium. Moreover, the background count rate from the sea bed differed considerably from place to place; the average background is about 200 counts per minute, but at some points it rises to 500 per minute. Even before the actual measurements (with the *artificial* tracer) were begun, therefore, it was possible to draw certain conclusions concerning sand movements, viz. to deduce at what places a relatively high content of heavy minerals and coarse grains of sand are deposited.

For introducing the radioactive isotope into the appropriate quantity of carrier material, an installation was erected on a barge; a photograph of the installation is shown in *fig. 10* and particulars are given in the caption [10]). The prepared material is deposited at the planned position from a small craft by means of a specially designed container, provided with a shield, which is let down vertically on to the sea bed to prevent the discharged material from being dispersed by the current before it has had time to settle.

The activity detector (a scintillation counter), properly shielded, is mounted on a sled two metres long (*fig. 11*), which is towed over the sea bed by a flat-bottomed boat. The detector thereby maintains

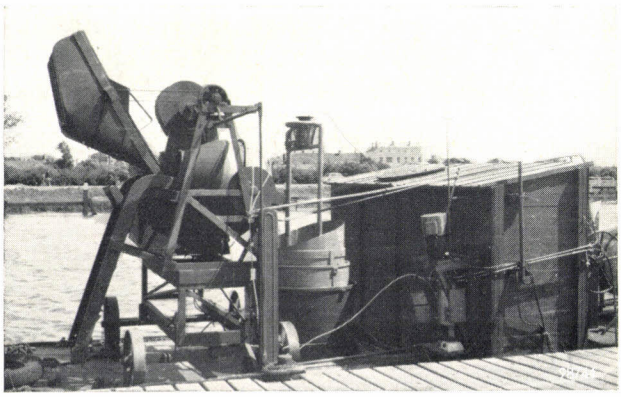


Fig. 10. Plant installed on a barge for preparing the tracer material [10]). An ampoule containing radioactive ^{46}Sc, prepared in a nuclear reactor, is broken open and the contents dissolved in nitric acid in a vessel (not visible in the photograph). The solution is neutralized and diluted with water, after which it is conveyed by compressed air to the mixer shown on the left, which contains the appropriate quantity of "Ionac C 50". After some hours of mixing the ^{46}Sc is distributed uniformly enough amongst the grains. The contents of the mixer are emptied into the container on the right of the mixer, in which the tracer material is conveyed to the dumping site. The large bunker on the extreme right is filled with sand to protect workers on the barge against radiation. The whole installation is remotely controlled from behind this bunker.

[10]) A description of the procedures in the field is given in quarterly report "Deltawerken" No. 6, Nov. 1958, pp. 22-28, from which figures 10-12 have been taken.

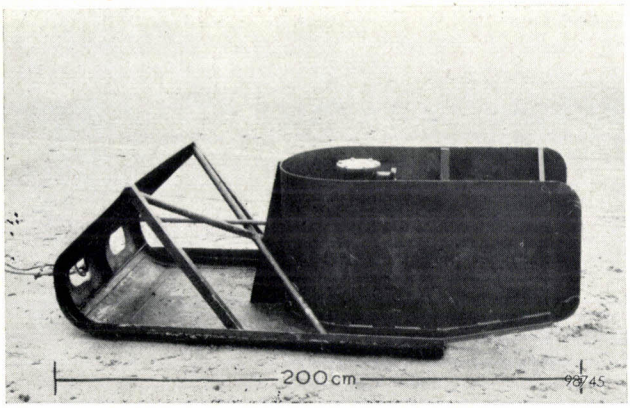


Fig. 11. Sled on which the scintillation counter and its shielding are mounted, and which is towed over the sea bed by a flat-bottomed boat. It is connected to the instruments on the boat by a 6-core cable.

automatically a constant distance of 20 cm above the surface of the sand (fig. 3). On board the boat are a count-rate meter and recording unit, together with a power supply unit which provides the 1700 to 2000 V constant potential for the detector. All this equipment is fed from a 220 V A.C. generator and is connected to the detector sled by a six-core cable.

The accuracy of the whole method is determined partly by the accuracy with which a fix is obtained on the position of the detector sled. Originally the boat was navigated as well as possible along a leading line defined by markers along the shores (*fig. 12*), and a fix of the boat's position on this line was obtained by means of a sextant. The Delta Authority now has at its disposal an apparatus for radio navigation, the "Decca Survey System", which was first used in 1944 during the Allied landings on the coast of Normandy. The system uses hyperbolic coordinates determined by three transmitters — situated at Rilland-Bath, at Sluis and at Schipluiden — and the navigator can fix his position in these coordinates from the readings of two indicating instruments on board the boat [11]). Within the estuaries this system has an accuracy of 3 to 4 metres, and outside them of 5 to 8 metres; with the old system the error in the fix could be as much as 20 metres. The position of the sled in relation to the boat is calculated from the length and sag of the towline and from the depth of the water, a correction being applied for the effect of the current.

In June 1959, large-scale measurements were started with the object of ascertaining the present littoral drift near the mouth of the Rotterdam Waterway. A total amount of roughly 300 kg of tracer material, having an activity of about 6 curies of ^{46}Sc, was deposited at three points marked *B*, *C* and *D* in fig. 2. It is expected that the results of these measurements, in conjunction with scale-model investigations, will allow conclusions to be drawn as to the effect which the building of Rotterdam's "Europoort" and the damming of the Haringvliet will have on conditions at the mouth of the Waterway.

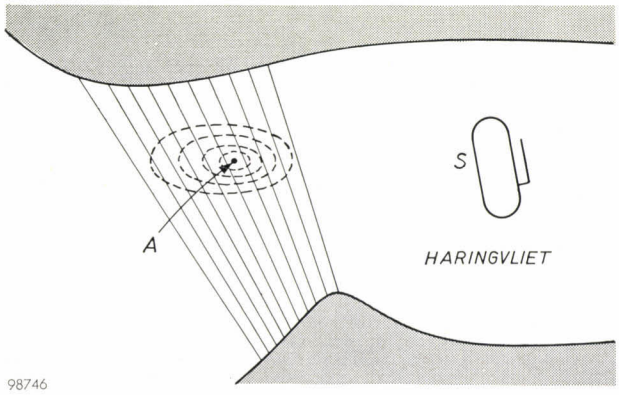


Fig. 12. Leading lines in the Haringvliet, defined by markers on the shores. The tracer material was dumped at *A*. The boat, while detecting and recording the radioactivity continuously, is steered as closely as possible along these lines so that with only one angular measurement a fix can be obtained on its position. The contours of equal activity, shown here by dotted lines, are determined by combining the results of all measurements taken at various times. This system of navigation has since been superseded by the "Decca Survey System", which is much more accurate.

[11]) Quarterly report "Deltawerken" No. 7, Feb. 1959, pp. 5-10.

Summary. The transport of sediment over sea and river beds by the action of currents and waves is of great importance to coastal and river works. During the execution of the Delta Project, which will radically alter the coastal currents off the west coast of the Netherlands, a careful watch must be kept on the movements of sand on the coast in order to prevent beach erosion and the silting-up of harbour entrances. A highly effective method of directly measuring littoral drift is to deposit a radioactively "labelled" quantity of sand at a point on the sea bed and thenceforth to monitor the activity in the surrounding area for a certain period. A rough calculation shows that very large amounts of radioactive materials (many curies) are required for this purpose. To investigate the problems involved, especially those relating to radiation protection, a preliminary experiment was made at the De Voorst hydrological laboratory by the Delta Authority in collaboration with the Philips-Duphar Isotope Laboratory. The tracer material consisted of glass beads made radioactive (isotope ^{24}Na) by irradiation in the Dutch-Norwegian nuclear reactor at Kjeller (Norway). In the actual measurements, now begun in the Haringvliet and at the mouth of the Rotterdam Waterway, use is made of ^{46}Sc, which has a more favourable half-life (85 days) than ^{24}Na and also emits a very suitable radiation. The deposited material must obviously have the same transport characteristics as sea sand. The experiment showed that this is by no means the case as far as the glass beads are concerned. The ^{46}Sc, however, can be bonded to a granular ion-exchange substance of the zeolite group (greensand, obtainable as "Ionac C 50"), which closely resembles North-Sea sand. To increase their hardness, the greensand grains are pre-fired at a high temperature.

AN EXPERIMENTAL DISC-SEAL TRIODE FOR 6000 Mc/s

by M. T. VLAARDINGERBROEK.

621.385.3.029.64

During the development work on the disc-seal triode type EC 157 [1]), it became clear that a frequency limit of 4000 Mc/s, and an output of 1.5 W at that frequency, did not constitute the utmost that could be achieved with triodes [2]). A triode capable of higher outputs — up to more than 10 W at 4000 Mc/s — has since been discussed in this journal [3]). With a metal-ceramic version of this valve in an oscillator circuit, it has been possible to step up the frequency to about 5000 Mc/s at an output of more than 10 W [4]). In the following a brief account will now be given of an experimental triode developed for amplifying signals in the 6000 Mc/s band (wavelength approximately 5 cm). This triode was tried out in an amplifier largely analogous to those made for the EC 157 and EC 59 valves. In this amplifier an average valve of this type delivers an output of 1.5 W with a gain of 7 dB at 6000 Mc/s for a bandwidth of 100 Mc/s.

The considerations involved in the design of a microwave triode have been dealt with on several occasions in this journal and elsewhere [5]). It will therefore be sufficient here to compare the new triode with the EC 157 type.

Theoretical considerations show that, to obtain a satisfactory gain with a triode, the transit time τ_{kg} of the electrons between cathode and grid must not be much greater than half the period of oscillation T of the electrical signal. The ratio τ_{kg}/T is given by the formula:

$$\frac{\tau_{kg}}{T} = 6.7 \times 10^{-10} f \left(\frac{d_{kg}}{J}\right)^{\frac{1}{3}}, \quad \ldots \quad (1)$$

where f is the frequency in c/s, d_{kg} the cathode-grid separation in cm and J the current density in A/cm^2. For the EC 157 ($d_{kg} = 40$ μ, cathode area 0.07 cm^2) the relevant values at $f = 4000$ Mc/s and at the usual anode current of $I_a = 60$ mA, are $J = 0.86$ A/cm^2 and $\tau_{kg}/T = 0.45$. If we wish to obtain the same value of τ_{kg}/T in a triode for 6000 Mc/s, equation (1) indicates that we can increase the current density J or shorten the distance d_{kg}, or both. The current density is limited to approximately the above-mentioned value by the power dissipation of the anode. On the other hand, d_{kg} cannot be indefinitely diminished because, owing to total-emission damping, the distance of the potential minimum from the cathode (about 5 μ at $I_a = 60$ mA) must remain small with respect to the cathode-grid separation. If this were no longer the case, the electrons that reverse before that minimum and return to the cathode would traverse a considerable part of the high-frequency field and convey R.F. energy to the cathode. Since this energy is not then used for modulating the electron current to the anode, this means a loss in gain. Experiments made on the EC 157 at 4000 Mc/s showed that the influence of the returning electrons, for $I_a > 30$ mA, is negligible [6]). These considerations led to the choice of $d_{kg} = 25$ μ for the 6000 Mc/s triode. For the same cathode area and at $I_a = 60$ mA, we then have $\tau_{kg}/T = 0.57$, which is an acceptable value.

In order to design a resonant cavity around the valve that is still tunable at these higher frequencies, one might, with the same anode surface area, reduce the anode-grid capacitance by making the grid-anode separation d_{ga} longer. This, however, would mean increasing the anode voltage in order to draw the same anode current, which would involve difficulties in cooling, owing to the higher anode dissipation. Another possibility of maintaining the tuning facility is to make the dimensions of the *outer* walls of the cavity round the valve smaller than for the EC 157. For an amplifier designed on the latter principle (see below), triodes have been made in which $d_{ga} = 250$ μ, i.e. approximately the same spacing as in the type EC 157. At an anode D.C. voltage of 250 V this enables sufficient current to be drawn from the cathode (> 60 mA). The maximum tunable frequency for the amplifier employed is higher than 6500 Mc/s. For the same amplifier, valves have also been made in which $d_{ga} = 300$ μ, and these have proved to be tunable up to more than 7000 Mc/s;

---

[1]) The EC 157 triode differs from the old type, EC 57, in having a cathode of longer life. The EC 57 is described in G. Diemer, K. Rodenhuis and J. G. van Wijngaarden, Philips tech. Rev. **18**, 317-324, 1956/57.

[2]) J. G. van Wijngaarden, Nouvelles possibilités des triodes à disques scellés, Onde électrique **36**, 888-892, 1956.

[3]) V. V. Schwab and J. G. van Wijngaarden, The EC 59, a transmitting triode with 10 W output at 4000 Mc/s, Philips tech. Rev. **20**, 225-233, 1958/59 (No. 8).

[4]) E. Mentzel and H. Stietzel, A metal-ceramic disc-seal triode for frequencies up to 6000 Mc/s, Philips tech. Rev. **21**, 104-108, 1959/60 (No. 3).

[5]) See articles [1]), [2]), [3]), and also H. Groendijk, Microwave triodes, Proc. Instn. Electr. Engrs. **105 B**, Suppl. No. 10, 577-582, 1958.

[6]) M. T. Vlaardingerbroek, Measurement of the active admittances of a triode at 4 Gc/s, Proc. Instn. Electr. Engrs. **105 B**, Suppl. No. 10, 563-566, 1958.

they require a somewhat higher anode voltage for the same anode current.

The losses in the anode resonant cavity are largely due to the absorption of high-frequency energy by the glass ring between grid disc and anode. By using a special hard glass — anode, cathode disc and grid disc are of molybdenum, whose expansion coefficient corresponds to that of various hard glasses and which is moreover a good electrical conductor — and by making the ring as thin as practicable, it was possible to keep the losses down to 15 to 20% of the power output.

The method of constructing the 5 cm triode differs from that of the type EC 157. A means was sought to simplify the greatest problem in the construction of a microwave triode, viz. that of achieving with adequate precision the very small spacing between cathode and grid. This simplification was reached by pre-assembling the parts shown darkly shaded in *fig. 1*. This assembly consists of two rings *5* and *6* of molybdenum which support respectively the cathode *1* and the grid frame *3* carrying the grid *2*; these rings are fixed with a precise separation by three rods *7* of synthetic sapphire. Sapphire was chosen for its great strength, its ability to withstand high temperature and for the ease with which it can be bonded to metal. We shall now discuss the fabrication of this cathode-grid assembly.

After the rods *7* have been bonded to the rings *5* and *6*, the grid frame *3* carrying the grid wires *2* is screwed to the ring *6*. The resultant assembly is then placed in an aligning jig (*fig. 2*) with the plane of the grid wires *2* exactly perpendicular to the spindle *A* of the jig and with the grid electrically insulated from that spindle. The cathode *1* is clamped in a sleeve *B* set in a chuck *T* on the spindle, the sleeve being pushed between the tantalum foil cylinder *8* and the cathode. With the screw *S* the spindle *A* is now moved axially until the cathode surface just touches the grid wires. Because the grid is insulated from the cathode, this contact can be detected with a short-circuit meter. The cathode is then screwed back to give the required separation as indicated by the dial gauge *M*. The problem of

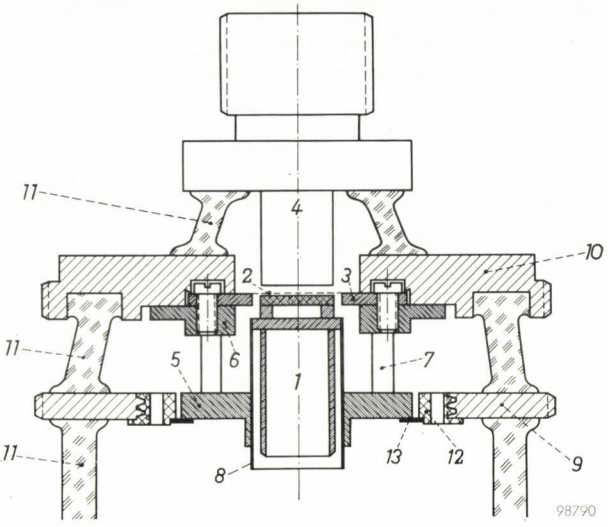


Fig. 1. Construction of the 5 cm triode. The darkly shaded section is pre-assembled. *1* cathode. *2* grid. *3* grid frame. *4* anode. *5* cathode ring (molybdenum) with welding collar (nickel). *6* grid ring (molybdenum). *7* sapphire rods joining the rings *5* and *6*. *8* tantalum foil cylinder of cathode, welded to the nickel collar of cathode ring. *9* cathode disc. *10* grid disc. *11* glass envelope rings. *12* closing ring with pump holes. *13* contact foil.

obtaining plane-parallelism between grid plane and cathode surface is relegated in this way to the construction of the jig. Backlash of the spindle is eliminated by the spring *V*.

When the cathode is at the desired distance, the tantalum foil *8* is welded to the cathode ring *5*. One difficulty here was that the last step, being a

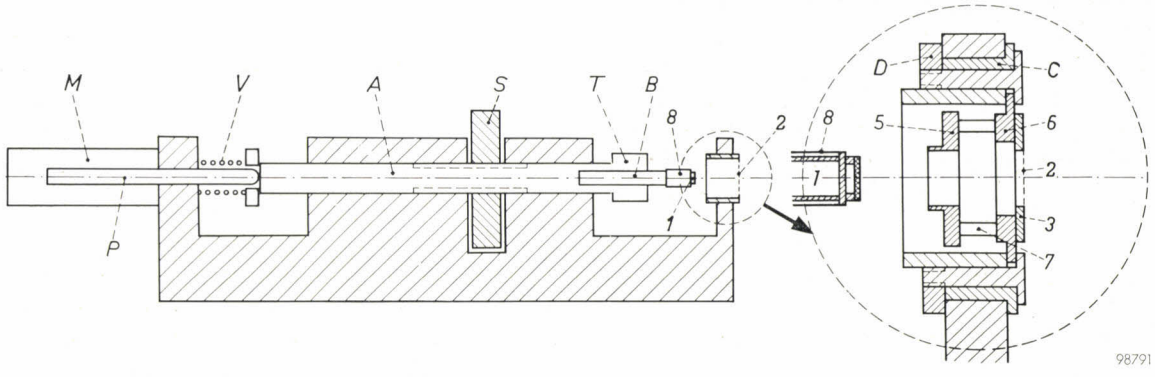


Fig. 2. Sketch of alignment jig for cathode-grid assembly. The valve components are numerically denoted as in fig. 1. *C* insulating ring of polytetrafluorethylene. *D* insulating ring of ebonite. The movable spindle *A* carries cathode *1* by means of the sleeve *B* mounted in the chuck *T*. *S* adjusting screw. *P* plunger of dial gauge *M*. *V* spring to prevent backlash.

welding operation, might cause plastic deformation of the tantalum foil and thus change the adjusted distance. It was found necessary to use nickel for the welding collar of the cathode ring, in which case only a light weld is needed to secure the tantalum foil. Plastic deformation is then negligible. Microscopic examination of a series of cathode-grid assemblies fabricated in this way showed that the cathode-grid spacing remained true to the adjusted spacing (approx. 47 μ) within 1 to 2 μ, and that the plane-parallelism was also good. The tantalum foil cylinder is shorter than in the EC 157, so that the change in length as the cathode heats up is smaller. The "hot" cathode-grid distance is about 25 μ.

The cathode-grid assembly can now be mounted in the outer structure of the valve (fig. 1). The grid-anode spacing is adjusted in this structure, after sealing the glass rings *11*, by turning on a lathe the anode *4* and the lower face of the grid disc *10* against which the grid ring *6* abuts. The cathode-grid assembly is fixed with three screws to the grid disc *10*. The annular opening needed between cathode ring *5* and disc *9* for this operation, is closed by a ring *12*, which is screwed into the cathode disc and makes contact with the cathode ring via a foil *13*. Pump holes are drilled in the ring *12*.

Fig. 3 shows a photograph of the aligning jig and the triode. The entire assembly takes place in a dust-free space.

The amplifier was designed, with the cooperation of J. P. M. Gieles, on the same principle as adopted for the EC 157 [7]) and the EC 59 [8]) types. For

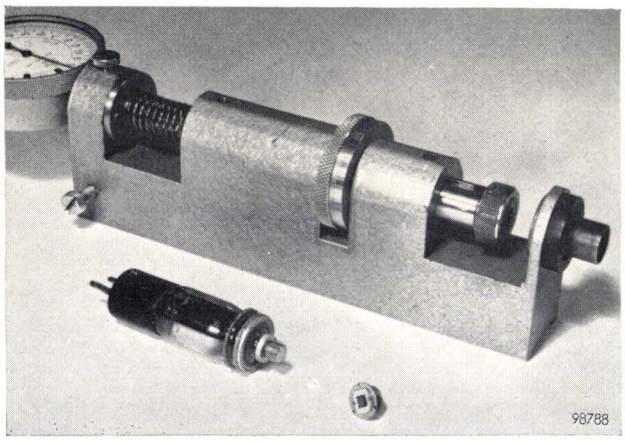


Fig. 3. Alignment jig, 5 cm triode and cathode-grid assembly (the latter with the grid to the fore).

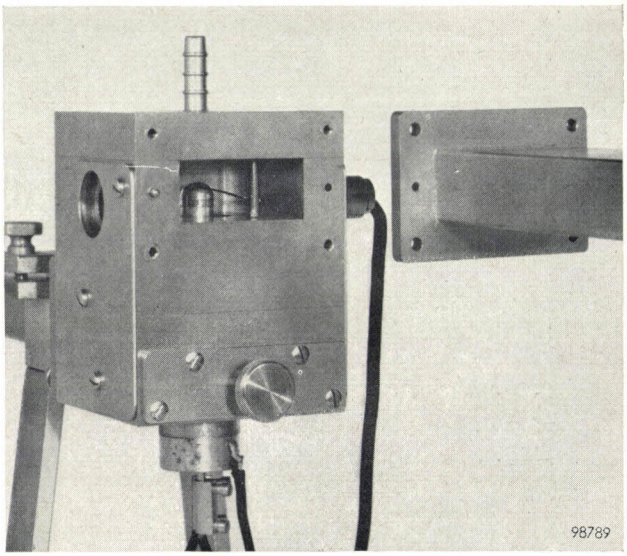


Fig. 4. Amplifier for the 5 cm triode, seen from the output side.

matching the amplifier to the input waveguide it was necessary to use a quarter-wave transformer. As in the amplifier for the EC 59, this was done by means of a ridge of length about a quarter wavelength inside the input waveguide [8]). Use is again made of additional reactive elements in order to correct for spread in the characteristics of individual valves. The bandwidth of the cathode circuit is greater than 500 Mc/s. At the output side the design differs from that for the EC 157 in that the inner conductor of the coaxial line is screwed to the anode. The principle, however, is the same. The amplifier can be seen in *fig. 4*.

Fig. 5 shows the static characteristics measured on an individual 5 cm triode. The transconductance is higher than 25 mA/V; the amplification factor is about 65.

The product of power gain G and bandwidth B of the anode resonant cavity is between 1000 and 2000 Mc/s for most of the valves tested (see *fig. 6*, measured on the same valve as fig. 5; the spread in $G \times B$ was found to be about $\pm 30\%$). It is noticeable that $G \times B$ is closely dependent on B (see caption to fig. 6). This is only partly explained by the fact that when B is small the losses in the anode cavity are considerably higher, amounting altogether to 20% at $B = 100$ Mc/s. The phenomenon is mainly attributable to internal feedback in the valve. This feedback is partly capacitative, partly inductive. The capacitative feedback is due to the anode-cathode capacitance (penetration — "Durchgriff" — of the alternating anode potential through the grid wires), and the inductive feedback may be represented by an inductive element in the grid

[7]) J. P. M. Gieles, Philips tech. Rev. **19**, 145, 1957/58.
[8]) J. P. M. Gieles and G. Andrieux, Philips tech. Rev. **21**, 41, 1959/60 (No. 2).

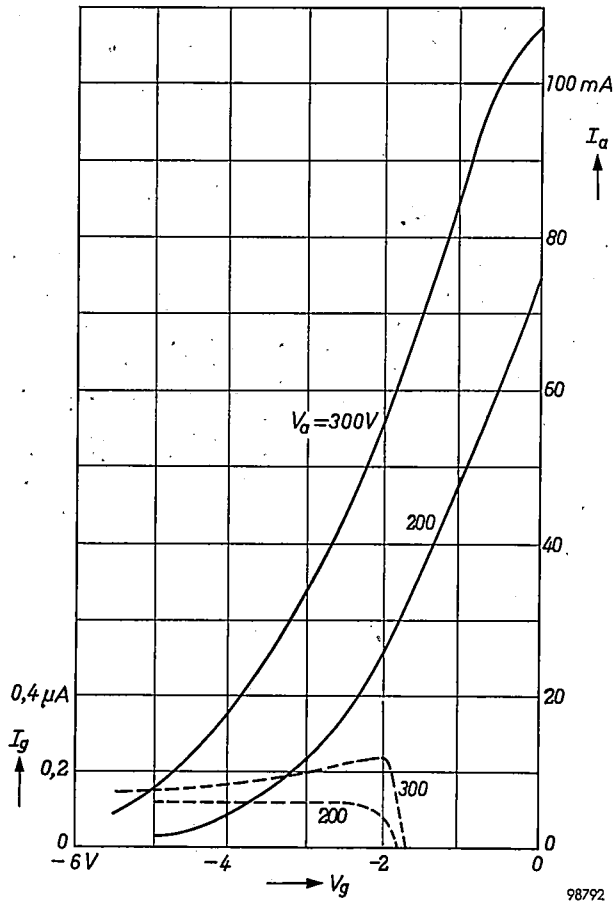


Fig. 5. D.C. characteristics of an average 5 cm triode.

lead [1]) [9]). Below the so-termed compensation frequency the capacitative feedback prevails, above this frequency the inductive. It can be shown that, below the compensation frequency, the product $G \times B$ must increase as B decreases. In the case of the EC 157, whose compensation frequency lies at approximately 5200 Mc/s, this is in fact found to be so at 4000 Mc/s. In the 5 cm valve triode, however, $G \times B$ is found to *decrease* with decreasing B (see fig. 6), from which it may be inferred that the operating frequency lies higher than the compensation frequency, i.e. that the inductive feedback is greater than the capacitive. If a grid with crossed wires were used, probably giving a smaller inductive feedback, it might be possible to increase the product $G \times B$ considerably for small B.

Fig. 7 shows the results of power output measurements made on the same valve to which figs. 5 and 6 refer. The curves show that an output P_0 of 2 W is possible. At $B = 100$ Mc/s an output of 1.5 W is achieved with a gain of 7 dB, which is 3 dB lower than with a small P_0. If B is increased P_0 decreases, the reason being that the anode impedance is then lower and P_0 is limited principally by the maximum current drive.

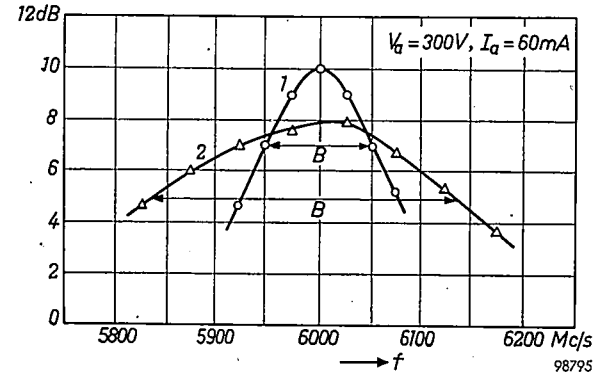


Fig. 6. The power gain in dB as a function of frequency at various bandwidths B, measured on the same valve as in fig. 5. The product $G \times B$ depends closely on B:
Curve 1 : $B = 100$ Mc/s, $G = 10$ dB, $G \times B = 1000$ Mc/s;
Curve 2 : $B = 300$ Mc/s, $G = 8$ dB, $G \times B = 1850$ Mc/s.

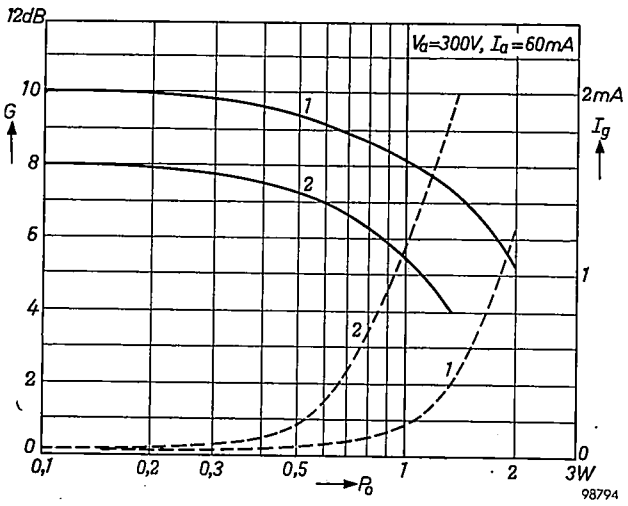


Fig. 7. Power gain G in decibels (solid curves) and grid current I_g (dashed) — which is required to remain within permissible limits — as functions of output power P_0 at bandwidths of 100 Mc/s (curves 1) and 300 Mc/s (curves 2), measured on the same valve as in figs. 5 and 6. The amplifier input is matched for a low P_0.

Suppose that the anode direct current is I_a and that the instantaneous value of the total anode current can be $2I_a$. In that case the r.m.s. value of the alternating current at the anode can reach $I_a/\sqrt{2}$. The anode impedance is $R_a = 1/(2\pi BC_a)$, where B is the bandwidth and C_a is the total capacitance of the anode circuit. The anode power is $(I_a/\sqrt{2})^2 R_a$; if B is 100 Mc/s, C_a is 1 pF and I_a is 60 mA, this represents about 3 W. The amplitude of the alternating anode voltage is then $I_a R_a \approx 100$ V, so that at a direct anode voltage of, say, 300 V there is comparatively little voltage modulation. The above considerations, which disregard feedback and losses, make it clear that it is the current drive that determines the relation between output power and gain.

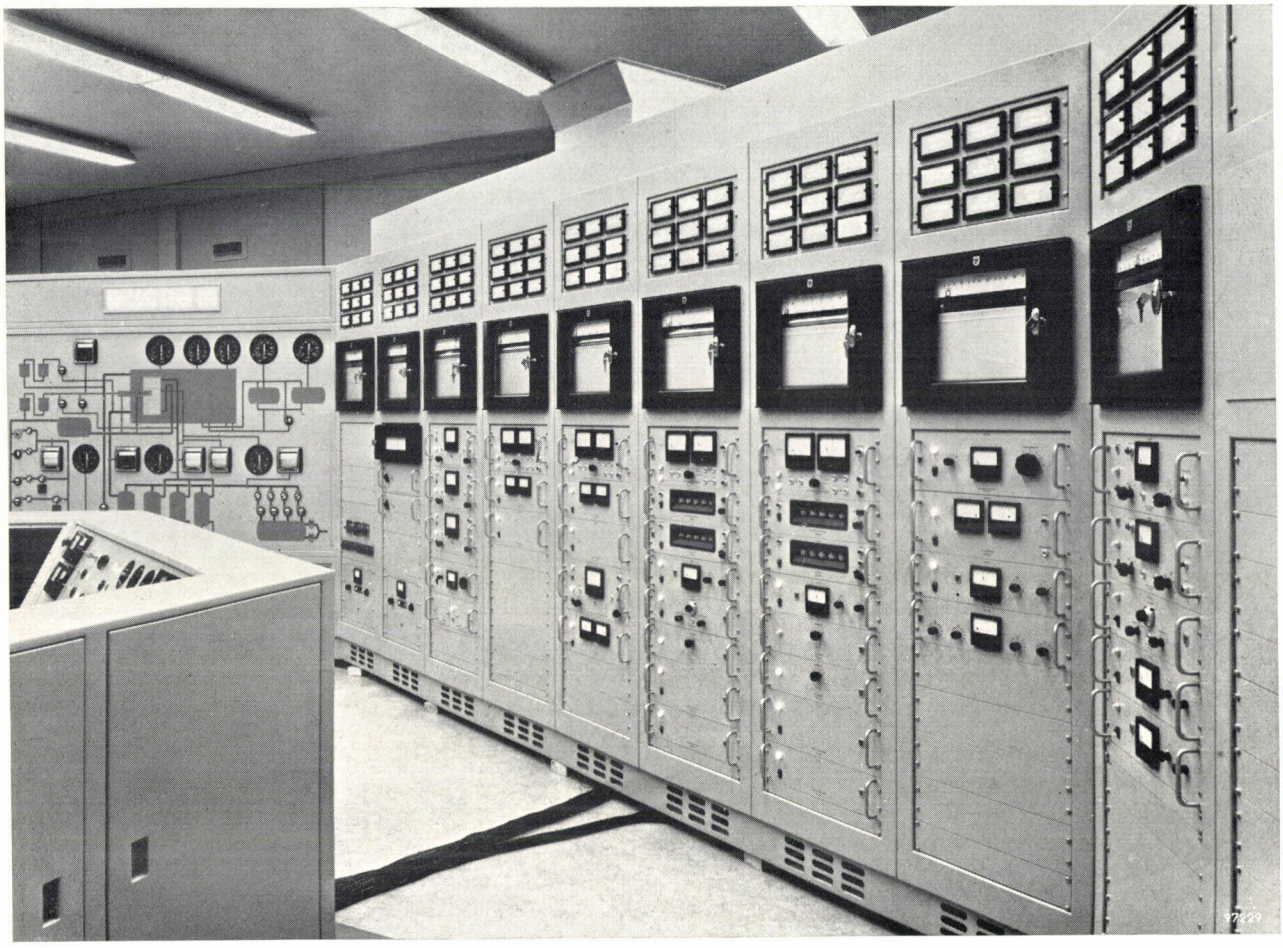
[9]) G. Diemer, Passive feedback admittance of disc-seal triodes, Philips Res. Repts. 5, 423-434, 1950. See also the article mentioned under [7]), page 152.

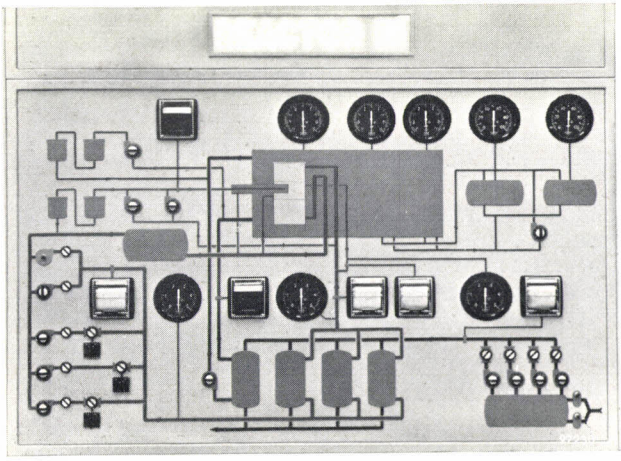
Finally, a few words about the valve as an oscillator. It was a simple matter to make an oscillator for 1.5 W output at 6000 Mc/s by feeding back part of the output to the input of the amplifier described. With the valves having a small grid-anode spacing this oscillator can reach a frequency of 6500 Mc/s. With the valves of larger grid-anode spacing a frequency of 7400 Mc/s has been achieved, but the power output is than lower. If the valve is incorporated in a suitable oscillator circuit, for example of the re-entrant type, even higher frequencies and outputs will certainly be possible, although this has not actually been demonstrated experimentally.

Experience gained in building this 5 cm triode, and the measurements made on it, justify the conviction that the upper limit of the frequency range at which triode amplifiers can operate with a sufficiently high $G \times B$ product has not yet been reached.

Summary. Brief description of an experimental disc-seal triode for 5 cm waves. The cathode-grid spacing is smaller than in the EC 157 and is achieved by pre-assembling these electrodes and adjusting the spacing in a precision jig. It amounts to about 25 μ with hot cathode. At a grid-anode spacing of 250 μ, made possible by reducing the outer dimensions of the anode resonant cavity, and at an anode voltage of 250 V, the highest tunable frequency of the amplifier designed round this valve is more than 6500 Mc/s; with 300 μ and a somewhat higher anode voltage it is more than 7000 Mc/s. With an average valve in this amplifier the product of power gain and bandwidth is 1000 to 2000 Mc/s. At 6000 Mc/s and a bandwidth of 100 Mc/s the power output is 1.5 W and the gain 7 dB. The losses amount to 15 to 20% of the output.

INSTRUMENTATION FOR THE PETTEN REACTOR





Near the village of Petten a 20 MW nuclear reactor is being built for the Netherlands Reactor Centre. The reactor is of the high neutron-flux type, and will be used for materials testing. Philips will supply the monitoring, control and safety equipment. The above photograph, taken at Eindhoven, shows the equipment in the final stage of assembly.

The equipment comprises instruments for measuring and controlling neutron flux, for safety monitoring, for measuring radiation density in the vicinity of the reactor, and for keeping a check on the radioactivity of cooling water and waste gases [1]). The instruments relating to the cooling circuit are on the graphic panel on the left (background of the photograph; a close-up is given in the smaller photograph). The circuit itself is represented schematically on the instrument panel.

([1] Installations designed for other types of reactor were described in Philips tech. Rev. **19**, 245-257, 1957/58, and Philips tech. Rev. **19**, 273-285, 1957/58 (swimming-pool reactor), and Philips tech. Rev. **21**, 109-152. 1959/60 (No. 4/5) (subcritical homogeneous suspension reactor).

METAL VACUUM EQUIPMENT

by N. WARMOLTZ and E. BOUWMEESTER. 533.599:621.791.856.3

It was not until the technique of arc-welding in rare-gas atmospheres was developed that it became really practicable to make vacuum-tight welds in the materials commonly used for vacuum apparatus — steel, non-magnetic chrome-nickel steel, aluminium, nickel and copper. In this welding technique, the electrical discharge between the workpiece and the electrode is shielded by a stream of a rare gas that displaces air from the vicinity of the hot metal, thus obviating oxidation.

The rare gases used — helium or argon in the United States, argon in Europe — are particularly suitable not only by virtue of their chemical inertness, but also because they are scarcely soluble in the metals named. If a gas is used that does dissolve easily in the molten metal, it escapes when the metal is setting, and pores or cracks may form in consequence. For example, hydrogen, which is fairly suitable for the welding of iron and milt steel, is quite unsuitable for welding alloys containing nickel, such as stainless steel [1]).

The possibility of making vacuum-tight welds in the metals normally employed in vacuum technique opens the way to the construction of leak-free metal components for vacuum systems. Such components can be assembled into systems which, besides being robust and highly vacuum-tight, can be efficiently degassed. Here we shall describe a number of such components, all of which have been welded in argon, and also a complete system which has been built up from them and which is suitable for a wide range of applications.

Of course the advantages just mentioned can be obtained by brazing as well as by welding. However, chrome-nickel steel — a material commonly used for vacuum equipment — cannot be brazed unless a firm layer of some other material has first been applied; moreover, the brazing of large and complicated workpieces is by no means an easy matter. Hence welding is generally to be preferred to brazing, and often enough it is the only practicable method [2]).

Our first example of an all-metal component is the cold trap shown in section in *fig. 1*. The trap has a particularly low flow resistance (high conductance). A cold trap of the same dimensions could of course be made of glass, but it would be fragile, and the connection to the vacuum system would give rise to some difficulty. A thick mantle of silica aerogel surrounds the vacuum chamber proper to provide thermal insulation. In this way the consumption of liquid air is reduced to less than 2 l in 24 hours. The trap is particularly suitable for mounting above the inlet of a mercury diffusion pump. The union between trap and pump will of course depend on the

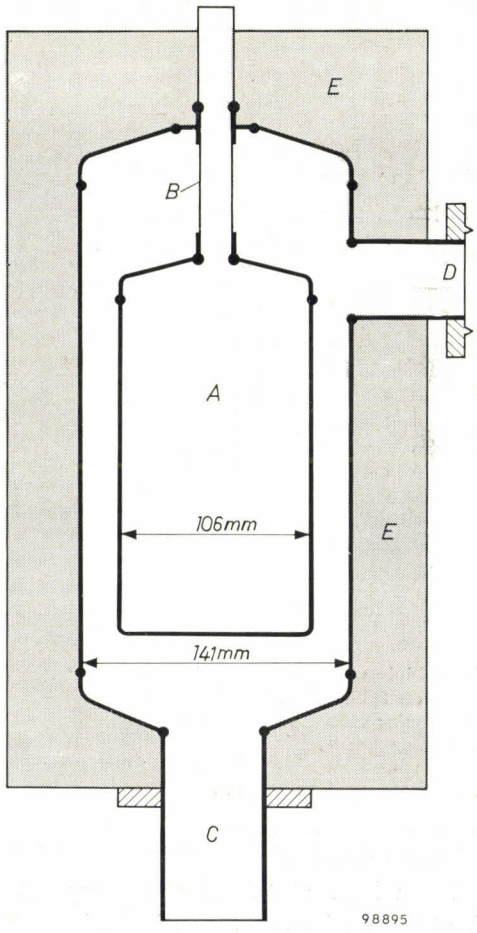


Fig. 1. Sectional diagram of an all-metal cold trap in stainless steel (18% Cr, 8% Ni, 74% Fe). Container *A*, holding about 1.5 l of liquid air, is suspended from tube *B*, which has a very thin wall in order to reduce its thermal conductivity. The trap is designed to fit on to a mercury diffusion pump, the connection being at *C*; at the temperature of liquid air (−183 °C) the trap has a conductance of 18.7 l/s for nitrogen. By allowing the trap to warm up slightly once every 7 to 10 days — during weekends for example — the mercury that has deposited on the outside wall of *A* has a chance to melt and to drip back into the pump of its own accord. The trap can be connected up to the vacuum system at *D*. Space *E* is filled with silica-aerogel insulation. Dots on the lines indicate argon-welded joints.

[1]) Cf. R. A. Weinman and I. Langmuir, General Electric Review **29**, 160, 1929.

[2]) A detailed account of the welding and brazing techniques employed in the manufacture of vacuum equipment may be found in W. Espe, Vakuumtechnik **4**, 51, 1955; argon arc-welding is described by N. E. Anderson in Welding J. **28**, 222, 1949.

design of the latter, and it is not therefore shown in the figure. The drawing does however, show (at *D*) the flange whereby the trap is attached to the rest of the system. All the components we shall go on to describe have flanges of this type, and it will therefore be as well to deal with them separately.

Fig. 2a is a more detailed drawing of a pair of these flanges. The flange on tube *1* has a rather blunt circular ridge *2*, immediately opposite a shallow circular valley in the flange of the other tube. A packing ring *3* made of fairly soft copper about 0.3 mm thick is placed between the two. The surfaces of the flanges, which should preferably be of stainless steel, must be highly polished: no scratches must be visible when they are examined under a glass giving tenfold magnification. Requirements for the copper packing need not be so strict: electrolytic copper, which does not have to be oxygen-free, is quite satisfactory. For tube diameters between 4 and 8 cm the flanges must be about 10 mm thick. If there is a glass-metal seal near the flange, it is advisable to give the flange a thickness between 15 and 20 mm, as otherwise the glass may break when the nuts of bolts *4* are tightened.

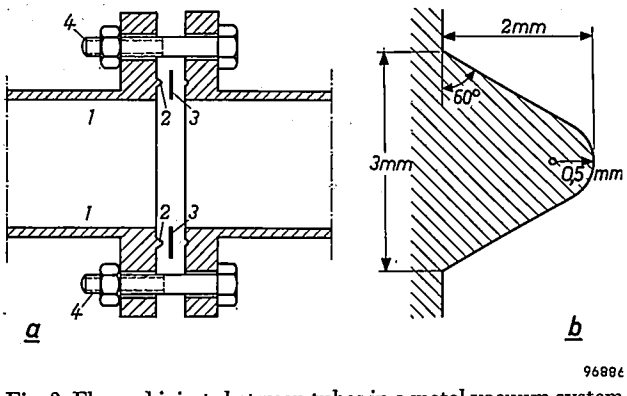


Fig. 2. Flanged joi nts between tubes in a metal vacuum system *a*) The flange on tube *1* bears a circular ridge *2* immediately opposite a circular score in the flange of the other tube. The packing *3* is of copper and is about 0,3 mm thick. The flanges are drawn together by a ring of bolts with a spacing of 2 to 3 cm. *b*) Section through the ridge *2*.

Experiments have shown that the joint just described can be maintained at a temperature of 500° C for 48 hours (e.g. for degassing) without undesirable effects. Nor do the nuts have to be tightened anew once the joint has cooled down. In view of the general need for degassing by heating, it is preferable to use copper as packing material rather than lead, which melts at a fairly low temperature, or aluminium, which flows and thereby necessitates retightening of the nuts.

The alternative method of connection illustrated in *fig. 3* can be employed for small tubes (diameter down to ~ 1 cm). Here the copper packing ring is clamped by the tube ends themselves, instead of by the flanges. Vacuum-tight welds are not therefore required.

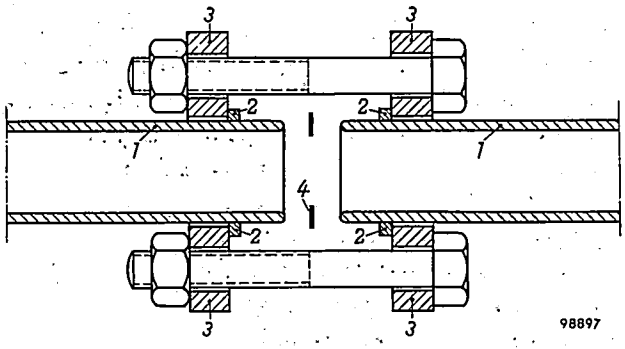


Fig. 3. Vacuum-tight joint directly between small tubes. The tubes *1* to be connected have their edges rounded off; *2* flanges welded on to tubes; *3* loose rings holed to take tiebolts; *4* copper packing ring.

We shall now discuss three vacuum valves, shown in *fig. 4*. The smallest is entirely of metal, while the two others embody a sealing ring made of "Teflon". When fully open, the largest valve has a conductance of 28 l/s for nitrogen, and hence it is very suitable as a pump isolating valve, for example. A cross-section of the valve is shown in *fig. 5*. The vacuum chamber is formed by the valve body *A* and the bellows *B*, both of which are made of stainless steel. The disc *D* moves on two guide rods *C* inside the valve body. The underside of the disc has a circular recess, rectangular in section and containing the Teflon sealing ring *E*. When handwheel *G* is turned, *D* moves downward and ring *E* is pressed on to the valve seating *F*. The Teflon ring, and the metal surfaces with which it comes into contact, are highly polished. An unsatisfactory seal in a valve of this type can easily be remedied by heating the closed valve for a short period, and so causing the Teflon to flow slightly. When open, the valve can be heated for a lengthy period to a maximum temperature of about 300 °C without suffering damage.

The design of the centre valve in fig. 4 is very similar to that of the one just described, closure again being effected by means of a Teflon ring. However, its conductance is lower (0.25 l/s). This valve is employed in gas inlet systems and in pipes to gauges.

Experience has shown that valves of either type hardly exhibit any wear or leakage after several years of service.

The smallest valve in fig. 4 (on the right) has an even lower conductance (0.1 l/s), but it has the advantage of being constructed entirely of metal; it can therefore be heated to a higher temperature than the two valves already described. It is not only

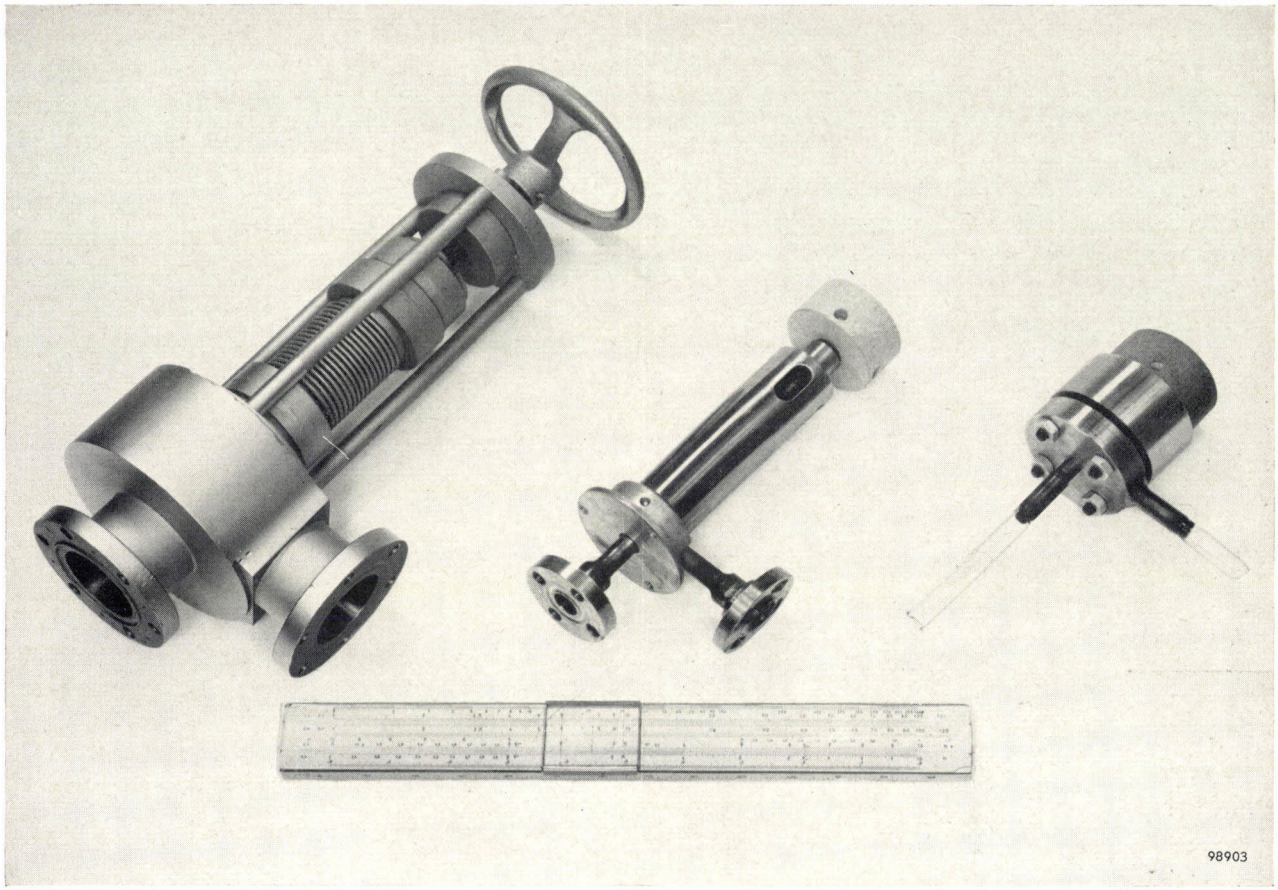
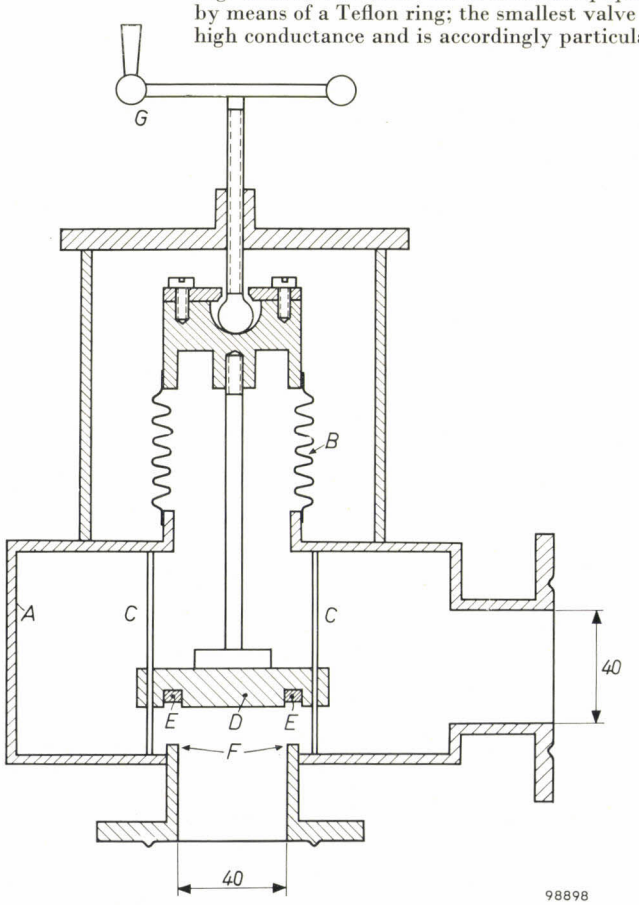


Fig. 4. Metal valves for use in vacuum equipment. In the two larger valves the seal is made by means of a Teflon ring; the smallest valve is entirely of metal. The largest has a very high conductance and is accordingly particularly suitable for use as a pump isolator.



in the absence of a Teflon seal that the design of the smallest valve differs from that of the other two, for a diaphragm (*A*, *fig. 6*) takes the place of the bellows. When new, these small valves have a leakage rate of less than 10^{-11} cm^3/s. After they have been heated once or twice, the leakage rate becomes about 10^{-8} cm^3/s. This low leakage makes the valve particularly suitable for use in gas inlet systems. Valves for such systems have to meet stringent requirements. While other kinds of vacuum valve normally only have to separate high from partial vacua, it is often necessary for gas inlet valves to seal off a storage chamber (at a pressure of 1 atm, say) from a high vacuum [3]).

[3]) These low leakage rates are measured, and the vacuum-tightness of components (and of complete systems) is investigated, with the aid of a small mass spectrometer used as a leakage detector. The lowest leakage rates that can be detected thus are of the order of 10^{-12} cm^3/s. A description of an instrument of this kind may be found in N. Warmoltz and H. A. M. de Grefte, Le Vide **12**, 202, 1957.

Fig. 5. Sectional diagram of large metal valve with a Teflon seal. *A* valve body. *B* bellows. *C* guide rods for disc *D*. *E* Teflon ring. *F* seating on which Teflon ring is pressed. *G* handwheel on threaded spindle.

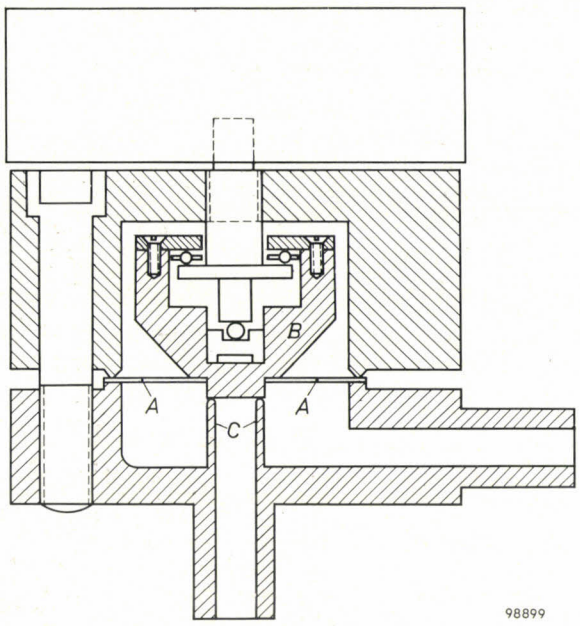


Fig. 6. Small all-metal valve. *A* diaphragm, *B* copper block that can be pressed against steel tube *C*.

above (fig. 1). The trap is connected via a large metal valve *D* having a Teflon seal (figs. 4 and 5) to a cylindrical chamber *E* on which are mounted an ionization gauge *J* (range 10^{-8} to 10^{-3} mm Hg) and a Penning gauge *P*. A flanged inlet is provided at the bottom of the cylinder, and to this the vessel or experimental apparatus *F* to be evacuated is attached. The pumping speed at the mouth of *B* is 30 l/s, and at the plane of the flange of *F* it has fallen off only to 10 l/s, as a result of using components of high gas conductance. Provided the system has been degassed sufficiently — for this purpose heating elements and thermal insulation are wrapped around *D* and *E*, and an oven surrounds *J* and *P* — it is possible to attain an ultimate pressure of 10^{-7} mm/Hg. If valve *D* is closed and evacuation continued with the ionization gauge alone, pressures down to 10^{-8} mm Hg can be reached [5]).

The purpose of the remaining part of the system is to enable the user to transfer to the vessel *F*

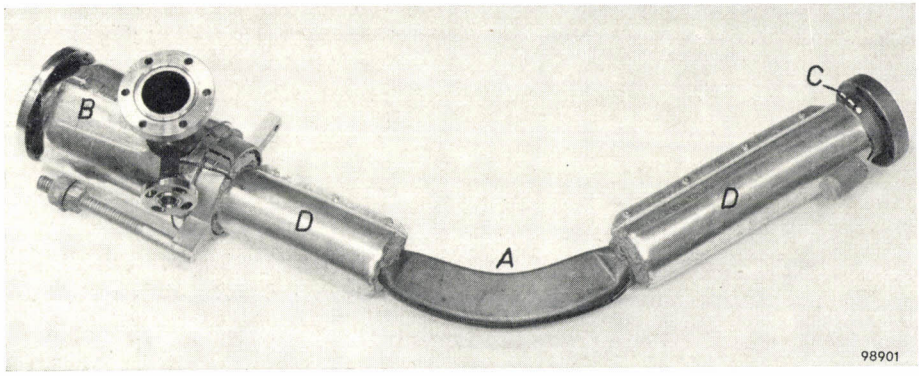


Fig. 7. Vacuum chamber for mass spectrometer. The curved central portion *A*, which is of copper, comes between the pole-pieces of a magnet when placed in the spectrometer. Connecting with the straight portions *D* are chambers *B* and *C* into which the ion source and the ion collector can be introduced. These chambers are made of non-magnetic stainless steel (25% Cr, 20% Ni). All parts were welded in argon except the steel-copper junction, which was brazed in a nitrogen atmosphere. The straight portions *D* visible in the photograph are actually heating mantles over the copper tubes.

Our last example of an argon-welded piece of vacuum equipment is the mass-spectrometer tube shown in *fig.* 7. Details of this component are given in the caption to the photograph.

The all-metal vacuum system mentioned earlier, which is suitable for a variety of applications, is shown diagrammatically in *fig.* 8. On the buffer tank *A*, which is exhausted by a rotary backing pump (not shown), is a mercury diffusion pump *B* [4]). Above that is a cold trap *C* of the type described quantities of one or other of the gases stored in the containers marked *Z*. These connect via small metal valves *K* (fig. 6) with pipe *H*. This pipe can be evacuated separately via pipe *N*. For this purpose valve G_1 is opened and valve G_2 is closed; both are of the type illustrated in the middle of fig. 4. The diaphragm gauge *M* [6]), which has a range between 10^{-5}

[4]) Mercury is preferred to oil for the reasons explained in A. Venema and M. Bandringa, The production and measurement of ultra-high vacua, Philips tech. Rev. **20,** 145-157, 1958/59 (No. 6), in particular p. 149.

[5]) See for example p. 146 of the article cited under [4]).

[6]) A description of this type of gauge is given in J. J. Opstelten and N. Warmoltz, Appl. sci. Res. **B 4**, 329, 1955. The Penning gauge is described in F. M. Penning and K. Nienhuis, Philips tech. Rev. **11,** 116, 1949/50. The ionization gauge is described in E. Bouwmeester and N. Warmoltz, Philips tech. Rev. **17,** 121, 1955/56).

and 10 mm Hg, can be used to measure the pressure of gas admitted into H from one of the containers Z. The whole of pipe H and its appendages, including the diaphragm gauge, are wound with heating elements to allow them to be degassed (heating up to 150 °C). In order to obviate damage, the electrical heating of the diffusion pump is automatically switched off if there should be any failure in the supply of cooling water to the pump, and also if the pressure in E should exceed a certain value. A photograph of the system appears in *fig. 9*.

With an all-metal vacuum system such as described here work proceeds more quickly and reliably than in the case with a glass system, and there is no loss of time on maintenance routines such as the greasing of cocks. Furthermore, thanks to the absence of grease and rubber, a system of this kind can be degassed by heating it in its entirety.

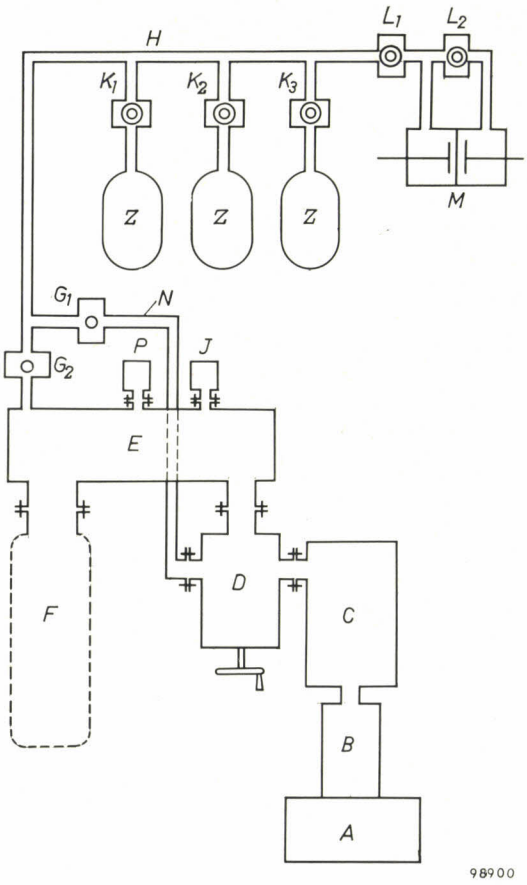


Fig. 8. Diagram of all-metal vacuum system with gas reservoirs. A buffer tank B mercury diffusion pump. C cold trap. D large metal valve with Teflon seal. E cylindrical chamber fitted with Penning gauge P, ionization gauge J and flanged connection to vessel F. G metal valves with Teflon seals. H tube for transfer of gas from reservoirs Z via small all-metal valves K; M diaphragm gauge with isolating valves L. The pipe H and its appendages can be separately evacuated via tube N.

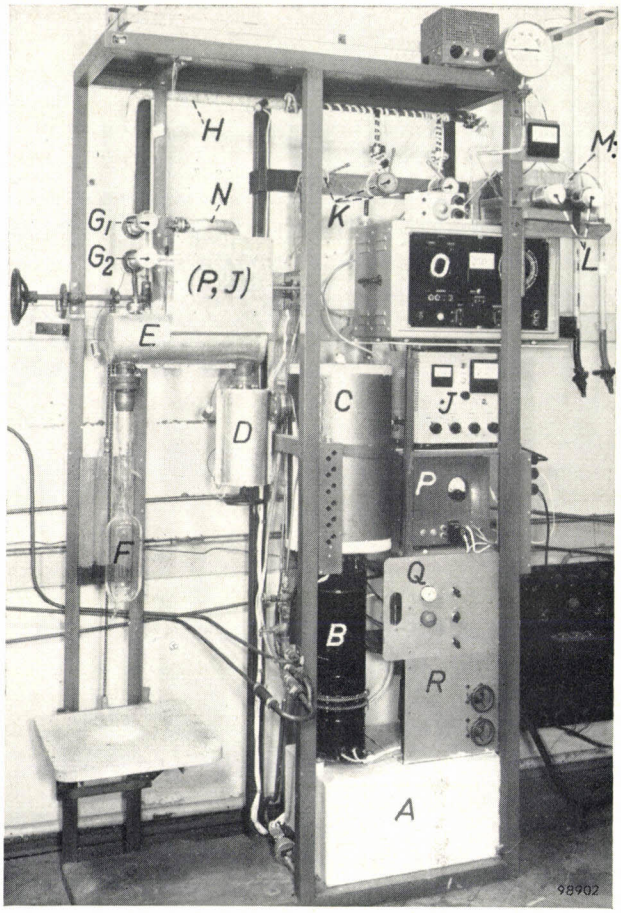


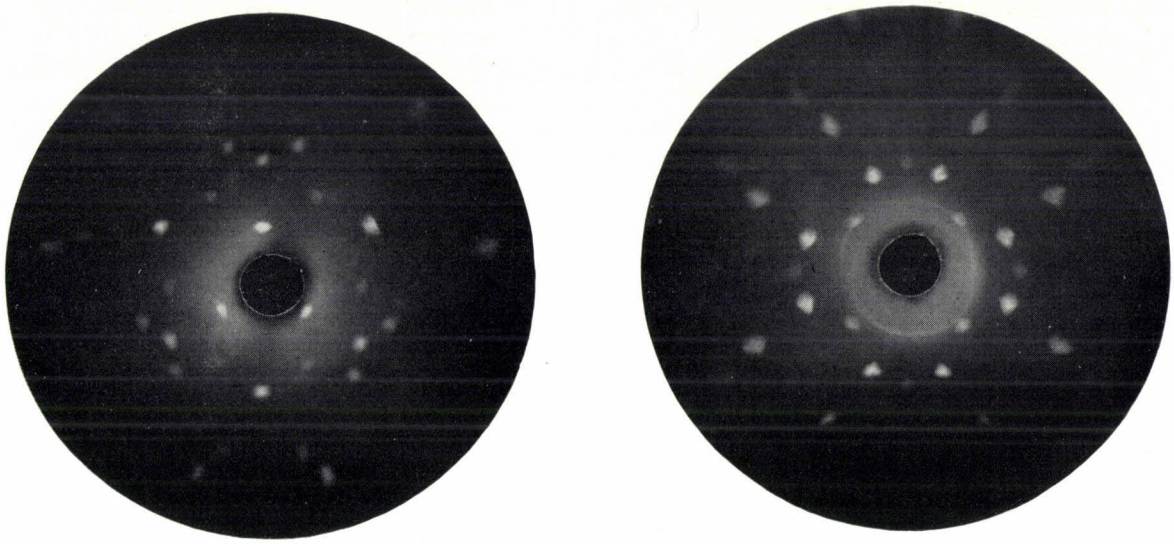
Fig. 9. Photograph of the vacuum system discussed in the text. The letters have the same meaning as in fig. 8; gauges P and J are behind the plate marked (P, J). The electronic circuits associated with the gauges, on the right of the photograph, are also marked P and J. O electronic circuit of diaphragm gauge M. Q switch panel, with fuse and pilot lamp. R panel carrying valve between tank A and rotary backing pump (not visible). All parts either have heating wires wrapped around them and, in some cases, insulation as well, or are located inside ovens (e.g. P and J).

Summary. The development of arc-welding in an argon atmosphere has opened the way to the construction of dependable vacuum-tight metal components for vacuum apparatus. The authors discuss components that have been produced with the aid of this welding technique, namely a cold trap and a vacuum valve, both of high conductance, two smaller valves, and a vacuum chamber for a mass spectrometer; they further describe the joints by which these components are coupled. The two larger valves are closed by a Teflon ring which presses against a metal seating; the smallest valve is entirely of metal. The packing material used for the joints is electrolytic copper. Finally, an account is given of an all-metal vacuum system which can be completely degassed and which is suitable for a large range of applications.

THE ORIENTATION OF DIAMONDS FOR TOOLS BY MEANS OF AN X-RAY IMAGE INTENSIFIER

621.9.027.2:621.386.8

Diamonds are nowadays used on a large scale in engineering industries. Because of its exceptional hardness, diamond is an ideal material for use in lathe bits, drills and grinders and for wire-drawing dies. The great demand for industrial diamonds and the comparative rarity of the mineral has led to the re-opening in recent years of closed mines and has also stimulated research in connection with the need to make the most economical and efficient use of the available stones [1]).



a *b*

Fig. 1. Laue diagram of diamonds, the X-ray beam being parallel to one of the axes of symmetry.
a) Beam parallel to a three-fold axis, *b*) beam parallel to a four-fold axis.

One result of this research has been to show that a diamond has the longest useful life when it is given a very specific orientation in the tool. As is usual in a single crystal, many physical properties are not the same in all directions. The resistance of a diamond to wear, for example, shows a pronounced anisotropy. Accordingly, if the diamond is so positioned that the direction in which the workpiece moves along the tool coincides with the direction in which the diamond shows the greatest resistance to wear, the wear it suffers can be reduced to a minimum. Sometimes the life of a tool is not so much threatened by wear as by cleavage. In that case a different orientation is called for.

The directions in which the physical properties of a single crystal show maximum or minimum values are of course related to the structure of the crystal. To be able to set a diamond we must know the orientation of its space lattice. This can be seen from the outward appearance of some diamonds, and such stones can be set in tools accurately enough without special aids. For orienting the others the method of X-ray diffraction can best be employed. As is well known, the space lattice of a crystal can be found by directing a narrow parallel beam of X-rays upon the crystal and examining the directions in which the beam is diffracted. The spot pattern which the diffracted beams produce on a photographic plate is the familiar Laue diagram. From this it is essentially a simple matter to determine the axes of symmetry of a crystal; if an axis of n-fold symmetry is parallel with the X-ray beam, the spot pattern also shows n-fold symmetry (*fig. 1*).

The diffracted beams are so weak that a Laue diagram cannot be made directly visible by directing the beams on to a fluorescent screen. Until recently, the X-ray-diffraction method of orienting a diamond — compared with other methods, by far the most accurate — was therefore a time-consuming process; after every alteration in position, a photographic exposure was necessary to ascertain whether the diamond was any nearer to the desired

[1]) See e.g. L. Schultink, H. L. Spier and A. J. van der Wagt, The wear of diamond dies, Philips tech. Rev. **16**, 91-97, 1954/55.

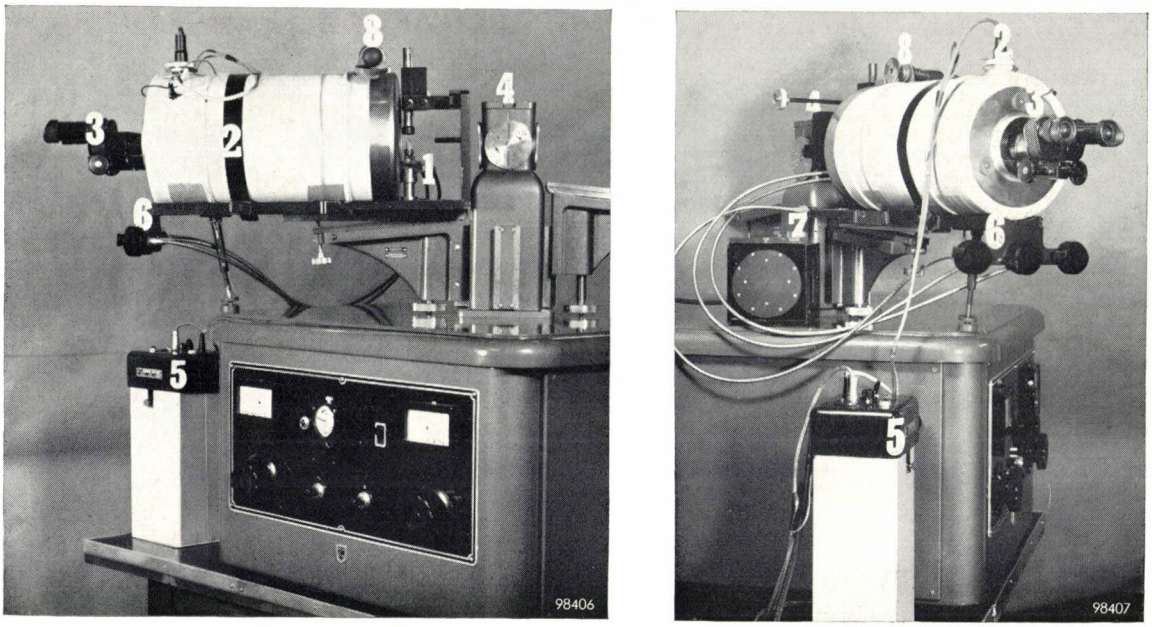


Fig. 2. Set-up for orienting diamonds, whereby the Laue diagram is observed with the aid of an X-ray image intensifier. *1* goniometer head on which the diamond is positioned; *2* X-ray image intensifier; *3* binocular viewer through which the intensified X-ray image is observed (the latter is 9 times smaller than the Laue diagram incident on the X-ray screen); *4* X-ray tube; *5* high-tension generator (25 kV) for supplying the image-intensifier tube; *6* controls with which the observer operates the goniometer head; *7* screen on which the desired diffraction pattern is optically projected for comparison purposes; the observer alters the position of the diamond until he observes in the image intensifier a pattern corresponding to the optical pattern on the comparison screen; *8* warning light, indicating that the X-ray tube is in operation.

orientation. This drawback of the X-ray-diffraction method has been removed by the advent of the X-ray image intensifier [2]), which permits direct visual observation of the Laue diagram. The process of orienting diamonds now takes only a few minutes and can be learnt by anyone in half a day.

Fig. 2 shows a set-up for orienting diamonds by the method described. The diamond is fixed to a goniometer head that can be remotely controlled by the observer behind the image intensifier. The observer alters the orientation of the diamond until he sees the desired spot pattern appear on the viewing screen. Since it is almost invariably possible (by inspection) to mount the diamond on the goniometer head in approximately the correct position, the orienting operation is, as said, relatively simple and rapid. Of course, the transfer of the diamond, thus oriented, to the holder in which it is to be set calls for a special technique, as developed, for example, in the Diamond Research Laboratory at Johannesburg, South-Africa [3]). The set-up shown in fig. 2 is located in this laboratory. A similar apparatus for orienting diamonds was built in Eindhoven some years ago [4]).

J. F. H. CUSTERS *) and A. J. van der WAGT.

[2]) See e.g. M. C. Teves and T. Tol, Philips tech. Rev. **14**, 33-43, 1952/53, and M. C. Teves et al., Philips tech. Rev. **17**, 69-97, 1955/56.

[3]) J. F. H. Custers, The orientation of diamonds for tools by means of an X-ray image intensifier tube, ASTE (Amer. Soc. Tool Engrs.) tech. Paper No. **104**, Coll. Papers **58**, Book I (1958). R. G. Weavind, C. J. Guykers and A. R. Roy, A rapid method of setting oriented diamonds in tools, ibid. Paper No. **103**.

[4]) See p. 95 of article [1]).

*) Diamond Research Laboratory, Johannesburg, South-Africa.

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 Laboratories, Eindhoven, Netherlands.

2655: H. C. Hamaker: Infusing statistics into industry (Bull. Inst. Int. Statistique **35**, 433-443, 1957).

The problems within industry to which statistics may be successfully applied are so many and varied that the solitary industrial statistician is not capable of coping with them all. One of his major tasks is therefore to teach others to apply statistical methods to their own problems. It is pointed out that the standard text books on statistics are not very suitable for this purpose. We should rather concentrate at once on the design of experiments, which is the most important chapter for the experimental scientist; and we should seek to explain the purposes and principles of a properly designed experiment and the corresponding analysis in simple terms without making too much use of a specific statistical jargon. How this can be achieved is illustrated by an example dealing with a foundry problem, where stacks, boxes within stacks, and moulds within boxes were the main factors entering into the analysis.

2656: H. de Lange Dzn.: Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light, I. Attenuation characteristics with white and colored light (J. Opt. Soc. Amer. **48**, 777-784, 1958, No. 11).

A generally known method for the dynamic investigation of any linear system is recalled to mind. Applied to the visual organ with sinusoidally modulated light, the dynamic nature of the system fovea → visual cortex is embodied in attenuation characteristics, by plotting the ratio output amplitude over input amplitude against frequency at constant mean luminance. This manner of investigation, first applied in previous papers with white light, is expanded over a greater part of the 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 of other authors.

2657: H. de Lange Dzn.: Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light, II. Phase shift in brightness and delay in color perception (J. Opt. Soc. Amer. **48**, 784-789, 1958, No. 11).

In this paper it is shown that the well-known residual brightness flicker just above the colour-flicker limit with heterochrome flicker photometry can be brought down to zero by introducing an external phase correction in one of two light beams sinusoidally modulated 100% in antiphase and simultaneously presented to the eye. The phase correction is found to be a function of luminance, colour difference, and frequency.

From the attenuation characteristic of the colour system it is found that the extra delay in colour perception at 595 mμ is consistent with a single integration process with a time constant of about 120 msec at high luminance; at low luminance a triple integration process seems to occur with the same time constant.

2658: R. Dijkstra and M. F. Lammers: On the determination of free phenol in phenol-formaldehyde resins (Rec. Trav. chim. Pays-Bas **77**, 933-934, 1958, No. 9/10).

The time necessary for the determination of free phenol in phenol-formaldehyde resins and moulding powders can be appreciably reduced by the addition of glycol to the resin during the steam distillation.

2659: G. Meijer: Influence of light on the elongation of gherkin seedlings (Acta bot. neerl. **7**, 614-620, 1958, No. 4).

Red and blue light have an inhibiting effect on the elongation of gherkin seedlings. A collaboration of blue and red light, given successively (B → R) or simultaneously (white light), however, has a much stronger inhibiting effect. This inhibiting effect can be neutralized by a subsequent irradiation with blue light; red or green light and darkness are ineffective in this respect.

2660: G. Meijer: The influence of light and of growth regulators on the elongation of gherkin seedlings (Acta bot. neerl. **7**, 621-626, 1958, No. 4).

The elongation of light-grown gherkin seedlings in blue light can also be obtained in red light in the presence of auxins. The elongating effect of blue

light on light-grown seedlings can be antagonized by anti-auxins. Light-grown seedlings transferred to darkness showed a toxic effect after a pre-treatment with blue or infra-red radiation. The same effect can be obtained by a pre-treatment with red light when auxins were added. Plants pre-treated with red light without an application of auxins did not show this typical effect.

2661: J. Kaashoek: Gradationsentzerrung im Farbfernsehen (Nachr.techn. Z. **11**, 515-518, 1958, No. 10). (Gradation correction in colour television; in German.)

Conventional methods of gamma correction in black-and-white television can be applied to colour television only to obtain a linear transfer characteristic. If these methods were used for gamma correction — often necessary for the transmission of transparencies and colour films — colour distortion would arise. This article discusses a new circuit which gives a controllable gamma correction dependent on the light absorption of the film, without colour distortion. The transfer characteristic of the system can be given any gamma value between 0.4 and 1. It is also possible to let the correction depend on the colours occurring in the picture, again without affecting the overall colour rendering.

2662: G. H. Jonker and W. Kwestroo: The ternary systems BaO-TiO_2-SnO_2 and BaO-TiO_2-ZrO_2 (J. Amer. Ceram. Soc. **41**, 390-394, 1958, No. 10).

An investigation of the ternary systems BaO-TiO_2-SnO_2 and BaO-TiO_2-ZrO_2 led to the discovery of two new compounds belonging to the system BaO-TiO_2., viz. $Ba_2Ti_5O_{12}$ and $Ba_2Ti_9O_{20}$. These compounds are stabilized by minute additions of SnO_2 or ZrO_2. The known compound $BaTi_2O_5$ can be obtained only from the molten phase and decomposes below 1300 °C into $Ba_2Ti_5O_{12}$ and $BaTiO_3$. In these systems no ternary compounds are found. The ternary phase diagrams can be divided into regions with high and low dielectric losses, which are in accordance with the phase relations. Tables with crystallographic data of the new compounds are included.

2663: F. A. Kröger: On the relation between non-stoichiometry and the formation of donor and acceptor centres in compounds (Phys. Chem. Solids **7**, 277-278, 1958, No. 2/3).

If the composition of a semiconductor compound deviates from the stoichiometric proportions (i.e. the composition defined by the simple chemical formula), donor and acceptor centres can occur. It is often assumed that one donor or acceptor centre occurs for every atom above the stoichiometric number. This is not necessarily the case, as is shown in this note.

2664: J. S. van Wieringen: La résonance paramagnétique comme moyen d'étude en recherche verrière (Verres et Réfractaires **12**, 256-260, 1958, No. 5). (Paramagnetic resonance as an aid in the study of glass; in French.)

The paper opens with a description of the principles of paramagnetic resonance (in which transitions between the energy levels of the elementary magnets of paramagnetic substances are measured). The apparatus used for such investigations is described. The author then examines why both paramagnetic glasses (as a result of their composition) and diamagnetic glasses irradiated by X-rays show paramagnetic resonance; in particular, investigation of the latter makes it possible to determine the structure of corresponding paramagnetic colour centres. A new method for the study of glasses is thus available, a method which has already given good results for paramagnetic salts, semiconductors, free organic radicals, etc.

2665: P. J. Papenhuyzen and P. Zijlstra: Une triode d'émission 10 kW pour télévision jusqu'à 220 Mc/s (Onde électr. **38**, 743-758, November 1958). (A 10 kW transmitting triode for television up to 220 Mc/s; in French.)

The first part of this paper concerns the design and properties of the transmitting valves TBW 6/20 (water-cooled anode) and TBL 6/20 (air-cooled anode). With a small output capacitance, large slope and large amplification factor, a bandwidth of 12 Mc/s is reached, as required for the French 819-line television system. The power amplification when used in grounded-grid configuration is large, viz. 9. The small output capacitance in conjunction with a small output self-inductance makes application in frequency band III (up to 216 Mc/s) possible. The valve is of the disc-seal type which facilitates mounting in a resonant cavity. The cathode has the form of a cage and is made of thoriated tungsten wire (mesh cathode). The second part of the paper describes a test transmitter in which the above two tubes are separately tested. The frequency of this transmitter is 216 Mc/s. From the tube dimensions, the dimensions of the resonant cavity are calculated. The are in good agreement with the dimensions of the cavity actually used. See also Philips tech. Rev. **19**, 118-128, 1957/58.

2666: J. L. Meijering and C. J. M. Rooymans: On the olivine-spinel transition in the earth's mantle (Proc. Kon. Ned. Akad. Wet. **B 61**, 333-344, 1958, No. 5).

The density gradient of the earth's mantle appears from seismic data to be greater than that to be expected from its compressibility. Bernal has suggested that this is due to a transformation of the olivine $(Mg,Fe)_2SiO_4$ (a major constituent of the earth's mantle) to a more close-packed structure, e.g. the spinel structure. The treatment of the earth's mantle as a one-component silicate system (Vening Meinesz) is found to be unsatisfactory. A treatment on the basis of the binary system Mg_2SiO_4-Fe_2SiO_4, on the other hand, gives more reasonable results. It is shown that the agreement might be further improved if the presence of still other silicates (such as $MgSiO_3$) were taken into account.

2667: T. J. de Man, E. J. ten Ham, J. R. Roborgh and N. Zwiep: Utilization of vitamin A from a stabilized dry preparation: comparative experiments with chickens (Netherl. J. agricult. Sci. **6**, 237-244, 1958, No. 4).

The utilization of vitamin A (continuously administered in low doses) from a stabilized dry preparation and from fish liver oil has been compared in experiments with chickens; these preparations were administered by mixing through the feed and by direct dosing into the beaks of the animals, respectively. Two criteria have been used: *a*) storage of vitamin A in the livers, *b*) growth. Modifications for both methods, permitting a quantitative statement of the results, have been developed. In order to obtain reliable figures, it was necessary to exclude any influence of differences in stability of the preparations investigated. Therefore, in the experiments during which the preparations were administered through the feed, this mixing was performed daily. The results, obtained by means of both criteria, lead to the conclusion that, aside from the question of stability, 1 I.U. of vitamin A in the form of the dry preparation, mixed through the feed, has the same biological activity in chickens as 1.3 I.U. of vitamin A in the form of fish liver oil. An experiment in which the mixing took place only once (5 weeks before the start of the experiment) resulted in an "activity ratio" of about 4.

2668: C. Wansdronk: Het meten van richtingsdiagrammen en overgangsverschijnselen aan luidsprekers (T. Ned. Radiogenootschap **23**, 303-309, 1958, No. 6). (The measurement of polar diagrams and transient phenomena of loudspeakers; in Dutch.)

Apart from the measurement of frequency and phase characteristics of loudspeakers, it is important to know their directivity pattern and transient behaviour. This article contains a description of the apparatus for rapid measurement of the latter two properties by making them visible on the screen of an oscillograph.

2669: J. W. L. Köhler: Fysische aspecten bij de ontwikkeling van koudgaskoelmachines (Ingenieur **70**, W. 175-W. 178, 5 Dec. 1958). (Physical aspects of the development of the gas refrigerating machine; in Dutch.)

Many problems linked with the Stirling process prove to be of a rather fundamental nature. The reason is that, whereas normal refrigerating apparatus consists of rather isolated equipment connected in series, the Stirling process is performed in a single machine, which combines all functions; this gives rise to a number of difficult cross-effects. Some examples of this are given — for instance the regenerator, which is the most interesting but difficult element of the machine. The cycle is interesting as a fine illustration of a reversible cycle; the results of its study may be used on a broad front, e.g. also for the hot-gas engine, the gas refrigerating machine and the heat pump.

2670*: K. Compaan and Y. Haven: Diffusion of tracers through solids (Proc. Internat. Symp. on Transport processes in statistical mechanics, Brussels, August 27-31, 1956, pp. 414-418; Interscience Publishers, New York 1958).

English translation of an article on indirect diffusion in crystals, already published in French as a chapter in the book "La diffusion dans les métaux", Philips Technical Library, 1957. See No. **2515** of these Abstracts and the short notice on p. 332 of Philips tech. Rev. **19** (1957/58).

2671: J. A. Kok: On superconductivity (Physica **24**, 1045-1050, 1958, No. 12).

This article deals with a two-component theory of superconductivity. At temperatures below the normal transition point the superconductor is assumed to consist of two components. One of these is the normal metal, whereas the second has a slightly larger volume $V + \Delta V$. The equilibrium between the electrons in the two states may be possible via the lattice vibrations. Some speculations are made on the normal transition temperature and on the question of what metals will become superconductive.

2672: H. J. Oskam and H. M. Jongerius: Helium-neon bands (Physica **24**, 1092-1094, 1958, No. 12).

In 1931 Druyvesteyn observed bands in the spectrum of helium-neon mixtures, which he attributed to a compound of helium and neon. The existence of these bands has since been doubted by others. From recent investigations (see e.g. Nos. **R 349** and **R 350** of these Abstracts), however, the existence of a molecule HeNe has been demonstrated by measurements of quite a different nature. The authors have therefore repeated Druyvesteyn's experiments; the results confirm his observations.

2673: M. J. Koopmans: A microdispersing apparatus for the use of biological investigation of chemicals (Meded. Landbouwhogesch. Opzoekingsstat. Gent **23**, 831-836, 1958, No. 3/4).

In investigations of the biological action of chemicals not soluble in water it is often required to put small quantities (10-50 mg) of them in the form of aqueous suspensions or emulsions. This article describes a mill suitable for grinding such preparations. A small glass vessel conically ground inside has a glass stop identically ground on its outside, which rotates inside it. The stop is supported from underneath, so that there remains a small play between stop and vessel. The conical stop has a flat on one side; in the space so left between stop and vessel the preparation and the water are introduced. A number of these mills are driven by an electric motor. After about 10 minutes the maximum particle size is 10 μ or less, dependent on the nature of the substance. The mean particle size is usually less than 2 μ.

2674: J. Brug, C. A. de Bock, H. D. Moed and A. J. Klein: Antiviral action of derivatives of ω-aminoacetophenone (Brit. J. Pharmacol. Chemother. **13**, 404-410, 1958, No. 4).

Numerous structural analogues of 3-amino-4-hydroxy-ω-methylaminoacetophenone were tested for their effect on the multiplication of influenza virus (FM_1 strain) in embryonated eggs, infected via the allantoic cavity. Antiviral activity was found in ω-aminoacetophenones containing an amino and hydroxyl group in the aromatic nucleus in the *ortho* or *para* positions to each other. The most powerful antiviral activity was found in the series of ω-alkylamino-5-amino-2:4-dihydroxy-acetophenones. Derivatives of acetophenone with other substituents in the aromatic nucleus or in the aliphatic chain were without activity. *In vitro*, the analogues exert a virucidal action. No inhibition of virus multiplication occurred in embryonated eggs infected via the yolk sac, or in mice infected intranasally with FM_1 virus. The activity in the allantoic test could be explained by the virucidal action of the compounds on the virus present in the allantoic fluid. No satisfactory interpretation of the empirical relationship between chemical structure and antiviral activity could be found.

2675: M. T. Vlaardingerbroek: Measurement of the active admittances of a triode at 4 Gc/s (Proc. Instn. Electr. Engrs. **105 B**, Suppl. No. 10, 563-566, 1958).

The paper describes the method of measuring the active admittances of the EC 57, a 4 Gc/s microwave triode. The results prove to be in reasonable agreement with the transit-time theory if the current density is sufficiently high.

2676*: J. H. Spaa: Algunos procedimientos para la producción de neutrones rápidos y monoenergéticos (Ciclo de conferencias celebrado con motivo de la exposición "El átomo y sus aplicaciones pacíficas", pp. 155-168; Sindicato Nacional de Agua, Gas y Electricidad, Madrid 1958). (Some methods for the production of fast mono-energetic neutrons; in Spanish.)

Report of a lecture given in Madrid on 21 May 1958. After a review of nuclear reactions potentially suitable for producing fast mono-energetic neutrons, two equipments for this purpose built in the Philips Research Laboratories are described. The first is a neutron generator based on the D-D reaction and using a 1 MV high-tension generator built in a tank pressurized with nitrogen at 10 atm. The neutrons are generated on the edge of a rotating dish containing heavy ice; cooling is effected by a liquid-air supply. The second equipment is a sealed-off neutron tube using a Penning ion source. Deuterons are accelerated and bombard tritium atoms absorbed on a zirconium target.

2677*: J. H. Spaa: Algunos nuevos métodos para detectar la contaminación radiactiva del aire y del agua (Ciclo de conferencias celebrado con motivo de la exposición "El átomo y sus aplicaciones pacíficas", pp. 173-187; Sindicato Nacional de Agua, Gas y Electricidad, Madrid 1958). (Some new methods for the measurement of the radioactive contamination of air and water; in Spanish.)

Report of a lecture given in Madrid on 23 May 1958. After some general considerations of the prin-

ciples of "Health Physics", two equipments developed in the Philips Research Laboratories are described. With the one equipment, which uses a special compensation technique, it is possible to detect and measure radioactive contamination of the air above the natural radioactivity. The apparatus works automatically, taking air samples and making measurements step-wise. The other apparatus is for the measurement of the β-activity of samples of very low activity. It consists of a mica-window geiger tube mounted in the hollow anode of a second geiger tube which is connected in anti-coincidence with the first. The whole is surrounded by a radiation shield. By suitable choice of materials, the design of the inner tube and the presence of the outer guard tube, an extremely low background is obtained. In this way extremely small activities can be measured in quite a short time.

2678: G. Meijer: The influence of light quality on the photoperiodic response of Salvia occidentalis (Acta bot. neerl. **7**, 801-806, 1958, No. 5).

Salvia occidentalis is a tropical plant which normally blooms only when the daily light period is shorter than a certain critical time (short-day plant). Earlier investigations by the author (see Nos. **2547** and **2548** of these Abstracts) showed that the blue and the infra-red components of the light hinder blooming. The investigation described here is concerned with whether this effect can be neutralized by suitable irradiation with red light.

2679: H. J. G. Meyer: Infrared absorption by conduction electrons in germanium (Phys. Rev. **112**, 298-308, 1958, No. 2).

Infra-red absorption by mobile charge carriers in semiconductors is governed by the band structure and the various scattering mechanisms. Measurement of this absorption therefore yields information on the band structure and the scattering. In this paper a quantum-mechanical theory is developed in which the multiple valley structure of the conduction band is taken into account. Using scattering theory in which a deformation potential is introduced, a fairly exact calculation can be made for all wavelengths and temperatures of interest. The result for the absorption due to normal vibrations of the acoustic branch contains two deformation-potential constants. Filling in the known values of these constants one finds, however, that the absorption coefficient at $\lambda = 10^{-3}$ cm and a temperature of 78 °K is a factor 5 too low with respect to the experimental value of Fan and Spitzer. This discrepancy can be eliminated by assuming that the normal vibrations of the optical branch also have an appreciable effect on the absorption. An estimate can then be made of the interaction between a conduction electron and these optical normal vibrations. With a simpler model an estimate is also made of the effect of impurity-centre scattering on the absorption coefficient. With the aid of this estimate, concentration, wavelength and temperature limits are given within which impurity-centre scattering can be neglected. The general limits of validity of the theory are given.

2680: H. Groendijk: Microwave triodes (Proc. Instn. Electr. Engrs. **105 B**, Suppl. No. 10, 577-582, 1958).

The valves treated in the paper are triodes with plane electrodes intended to be used as amplifier valves, local oscillators and transmitting valves in repeaters working in the super-high-frequency region. The most important factor in their design is the product of power gain and bandwidth. It is shown that this product is chiefly determined by valve parameters. For a high value of the gain-bandwidth product, a high current density and a small cathode-grid distance are needed, resulting in a small electron transit time. A transit time of 1.3×10^{-10} sec has been obtained. Although the deviations of input conductance and transadmittance from their D.C. values — due to transit-time effects — are considerable, these deviations have only a small influence on the gain-bandwidth product. A table is given showing constructional and electrical data for a number of microwave triodes, of which one is in production. The others are in the experimental stage of development. The paper discusses the following properties of an amplifier with a microwave triode: (*a*) Gain-bandwidth product, 2 Gc/s. (*b*) Noise figure, 16 dB. (*c*) Variation of group-delay, 0.06 millimicrosec. (*d*) Amplitude-phase conversion, 0.8 deg/dB. The figures mentioned are obtained with an amplifier stage equipped with an EC 57, a triode for operation at 4 Gc/s.

VOLUME 21, 1959/60, No. 7 pp. 185-220 Published 7th June 1960

Philips Technical Review

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

SEGREGATION AND DISTRIBUTION OF IMPURITIES IN THE PREPARATION OF GERMANIUM AND SILICON *)

by J. GOORISSEN.

548.4:669.783:669.782

In connection with solid-state research and the production of circuit devices based on semiconductors, much work has been done in recent years on developing and perfecting methods of refining the substances involved and introducing controlled amounts of impurities. The article below describes two such methods (variants of the zone-melting technique) for preparing "doped" single crystals of germanium and silicon. These methods, evolved in the Philips Eindhoven laboratories, yield a product in which the concentration of the impurity is uniformly distributed.

In the fabrication of transistors and associated semiconductor devices it is necessary to be able to prepare the materials germanium and silicon with a high degree of purity, and also to "dope" them in a controlled way with a particular impurity. Further, the product is required in the form of a single crystal and must meet certain demands as regards physical perfection.

In all the various processes the product is obtained by the solidification of molten material. This involves the phenomenon, which has favourable as well as unfavourable aspects, that the impurity concentration in the solid at the freezing point differs from that in the liquid. If the concentrations are small, their distribution coefficient k_0 (i.e. the ratio of the concentration C_S in the solid to the concentration C_L in the liquid) has a constant value that usually differs very considerably from unity. For example, in the case of a solution of indium in germanium k_0 is only 1.4×10^{-3} (*fig. 1*).

In the following we shall first consider the consequences of this phenomenon when a dilute solution freezes progressively from a given point, i.e. when the solid-liquid interface moves through the charge. We shall see that, where $k_0 < 1$, a large fraction of the impurity gathers in the last part to solidify, so that the concentration in the remainder is reduced. This segregation process can therefore be used for purification. In the solidified material, however, the impurity concentration is not uniform, and from the point of view of doping, where the aim is to obtain a homogeneous final product, this is a troublesome effect. In the latter half of this article we shall discuss two doping methods developed in this laboratory, whereby special measures are taken to ensure

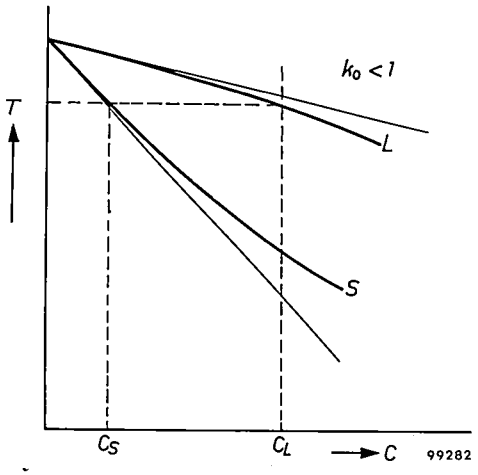


Fig. 1. In a two-component system the concentrations at a given temperature in the liquid and solid phases are not equal. When a liquid of concentration $C = C_L$ is cooled, crystallization occurs as soon as the temperature drops below that corresponding to C_L on the liquidus curve L. The concentration C_S in the resultant solid is the abscissa value of the solidus curve S at that temperature. For small concentrations S and L may be replaced by straight lines, and the ratio C_S/C_L is equal to the distribution coefficient k_0. If the two lines drop with increasing concentration, then $k_0 < 1$; if they both rise, then $k_0 > 1$.

*) This article is based on a lecture given 18 April 1959 at Moll (Belgium) before the Int. symp. on pure metals and semiconductors. The text of the lecture will shortly be published in the Vlaams Chemisch Weekblad.

that the impurity concentration is constant throughout a large part of the volume of the product.

Segregation and distribution of impurities in the normal solidification process; crystal pulling

In considering exactly what happens at a moving solidification boundary we shall suppose that the material is contained in an oblong boat in a furnace and is wholly molten. The boat is now slowly withdrawn from the furnace, so that the melt solidifies at one end. In relation to the furnace the solid-liquid interface remains at the same position, but in relation to the boat it moves from one end to the other. We assume that the solid-liquid interface is perpendicular to the length of the boat, which allows us to treat the problem as one-dimensional; the length of the boat represents the x axis. The origin is the place where the solidification begins. The x coordinate of the solid-liquid interface we shall call s (*fig. 2*). As long as the whole charge is

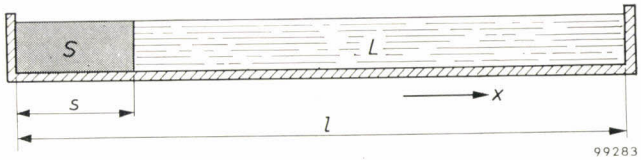


Fig. 2. Progressive solidification in a boat of length l. The x axis is the long axis of the boat. The solid-liquid interface is at the location $x = s$ and moves to the right.

molten, the concentration of the impurity is everywhere the same; let this value be C_i. In the material first to solidify, the concentration C_S is equal to $k_0 C_i$. For $k_0 \ll 1$, this solid part is thus appreciably purer than the liquid. The amount of impurity expelled from the volume which has frozen is now contained in the liquid. The further variation of C_S with increasing s depends on the manner in which this displaced amount of impurity is distributed over the bulk of the liquid. In principle, two limiting cases are conceivable: *a*) due to vigorous stirring, the amount of impurity taken up is immediately distributed uniformly throughout the bulk of the liquid, and *b*) there is no convection or stirring, so that the impurity is transported in the liquid solely by diffusion.

In the first case, for s fairly small relative to the length l of the boat, the concentration C_S in the solid increases only slowly with x; it has only double the $x = 0$ value at points $x \geqq l/2$.

In the other case, a concentration distribution arises that depends on the diffusion constant D and the rate f at which the solid-liquid interface moves, such that C_L *at* the interface is equal to C_i/k_0, and C_S is equal to C_i. This is illustrated in *fig. 3a*. Further from the interface, C_L still (in general) has the value C_i; the distance between the interface and the place where $C_L(x)$ differs from C_i by less than, say, 1% is denoted d in fig. 3*a*. When $D/f \ll l$, this "length" d of the diffusion region is for a long time smaller than the length $l - s$ of the liquid. Assuming, then, that for $x > s + d$ the liquid has exactly the concentration C_i, and since in the steady state as much impurity must enter and leave the diffusion region via the solid-liquid interface as enters and leaves it from the other side, evidently $C_S = C_i$. Except at the ends of the bar, C_S is thus equal to C_i in this case.

In reality, convection is not completely absent; there is, however, a boundary-layer of liquid at the advancing interface in which convection in the x-direction is negligible and in which the impurity can therefore be transported only by diffusion. If the circumstances — we shall turn to this point presently — are such that the thickness δ of this layer is less than the above-mentioned diffusion length d, the diffusion pattern will differ from that

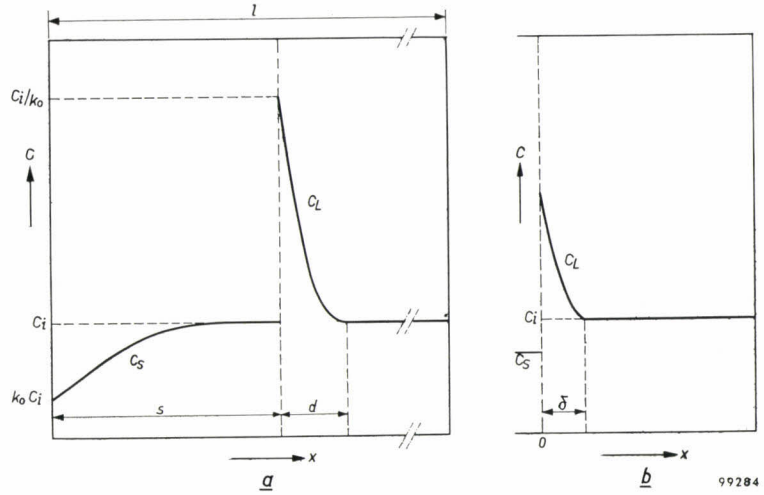


Fig. 3. *a*) If the transport of the impurity not taken up in the solidified material is due solely to diffusion, the diffusion process at the solid-liquid interface reaches a steady state as soon as C_L at the interface has risen to the value C_i/k_0.
b) The actual situation at the solid-liquid interface can be described by assuming the presence at this interface of a layer of thickness δ in which the impurity is transported in the x direction solely as a result of diffusion, whilst in the bulk of the melt convection occurs. When $\delta < d$ (fig. 3*a*), C_L at the solid-liquid interface does not attain the value C_i/k_0, and k ($= C_S/C_i$) is therefore smaller than 1. (In connection with the derivation of equation (1), the origin of the coordinate system is situated here in the interface.)

in the case without convection. The result is that C_S at the interface cannot reach the value C_i/k_0. The segregation coefficient k, which is equal to C_S divided by C_L measured at some distance from the solid-liquid interface, therefore has a value in this case between k_0 and unity (fig. 3*b*). It can be shown that this value is given by the formula [1]):

$$k = \frac{k_0}{k_0 + (1-k_0)\, e^{-f\delta/D}}. \quad \ldots \quad (1)$$

Taking a coordinate system whose origin is in the solid-liquid interface (and which thus moves with the advancing interface), we can write (cf. fig. 3*b*):

$$D\frac{d^2C}{dx^2} + f\frac{dC}{dx} = 0 \quad \ldots \ldots \quad (2)$$

(omitting the subscript L), the boundary conditions being:

$$f\{C(0) - C_S\} + D\left(\frac{dC}{dx}\right)_{x=0} = 0, \quad \ldots \quad (3)$$

$$C(x) = C_i \text{ for } x \geqq \delta. \quad \ldots \ldots \quad (4)$$

The solution is:

$$C(x) = C_i \frac{k_0 + (1-k_0)\, e^{-fx/D}}{k_0 + (1-k_0)\, e^{-f\delta/D}}. \quad \ldots \quad (5)$$

Putting $x = 0$, we find from this expression the concentration $C(0)$ at the interface, and multiplying this by k_0, we find the concentration C_S in the solid. Dividing the expression thus obtained by C_i, we finally arrive at the equation (1) for k.

If we substitute in (1) the value of k under given conditions, we can find the magnitude of δ from this expression. For antimony in germanium, δ is found to be of the order of 10^{-2} cm at an f value of 1.8 mm/min. (The case in which mass transport is due solely to diffusion can also be expressed by (2), (3) and (4): δ is then put equal to infinity in the second boundary condition (4). The solution — represented graphically by the curve in fig. 3*a* — is found from (5) by again putting $\delta = \infty$; the denominator of (5) then equals k_0. From this solution we can deduce that the distance d from the solid-liquid interface, where $C(x)$ differs from C_i by not more than 1%, is approximately 1.7 mm in the example of Sb in Ge.)

A plot of C_S versus x is given in *fig. 4* for the actual situation (curve *3*) as well as for the two hypothetical limiting cases. Curve *1* applies to the case where there is complete absence of diffusion but very vigorous stirring. Curve *2* applies to the case of no stirring. Except near the head of the sample, where the diffusion process has not yet reached the steady state, curve *3* shows the same general form as curve *1*. It can be seen that, although a substantial fraction of the impurity is displaced towards the last part to solidify, everywhere in the sample the value of C_S varies somewhat with x. This means that although the method of solidification described above is suitable for purposes of purification — and, as such, is comparable with the re-crystallization methods used in chemistry — it is less suitable as a method of doping.

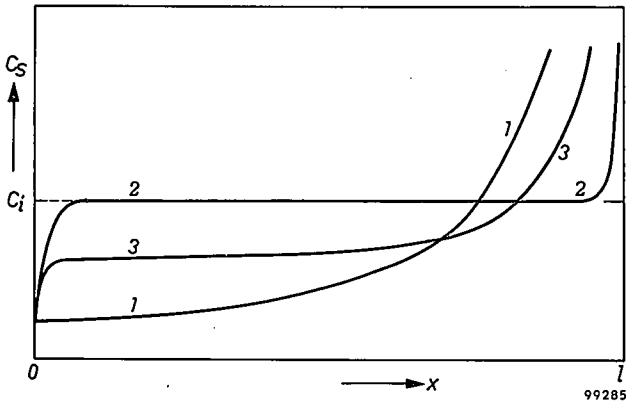


Fig. 4. The variation of C_S with x in ingots produced by the progressive solidification from one end of a melt having a concentration C_i. Curve *1*: the melt is vigorously stirred. A large portion of the ingot is highly purified, but C_S does not have a constant value. Curve *2*: no convection in the melt, the impurity being transported solely by diffusion. Over a considerable length of the ingot $C_S = C_i$, i.e. k is constant over this region and equal to unity. In reality these extremes do not occur, and C_S follows curve *3* ($k_0 < k < 1$). Except where x is very small, curve *3* has the same character as curve *1*.

Except at very small values of x, the variation of C_S with x is given for all possible situations by:

$$C_S(x) = kC_i(1 - x/l)^{k-1}. \quad \ldots \quad (6)$$

In the first limiting case mentioned above (fig. 4, curve *1*) we had $k = k_0$, and in the other case (curve *2*) we had $k = 1$. In reality, of course, the value of k lies somewhere between k_0 and 1.

Finally, it should be noted that $C_S(x)$ can be indirectly determined by measuring the local resistivity of the material. The relation between the two quantities is known: the conductivities of impure germanium and silicon are to the first approximation proportional to C_S.

The pulling of single crystals

A widely used method of making single crystals is the Czochralski technique [2]). A seed crystal attached to a rotating vertical shaft is lowered until it just touches the surface of a quantity of molten material. The temperature of the melt at that position must be just equal to the freezing point in order to prevent the seed crystal from melting away. The shaft with the seed crystal is now slowly raised, and at the same time the heat supplied to the melt is slightly reduced. As soon as a steady state is reached, a rod-shaped single crystal

[1]) J. A. Burton, R. C. Prim and W. P. Slichter, J. chem. Phys. **21**, 1987, 1953.

[2]) J. Czochralski, Z. phys. Chem. **92**, 219, 1917.

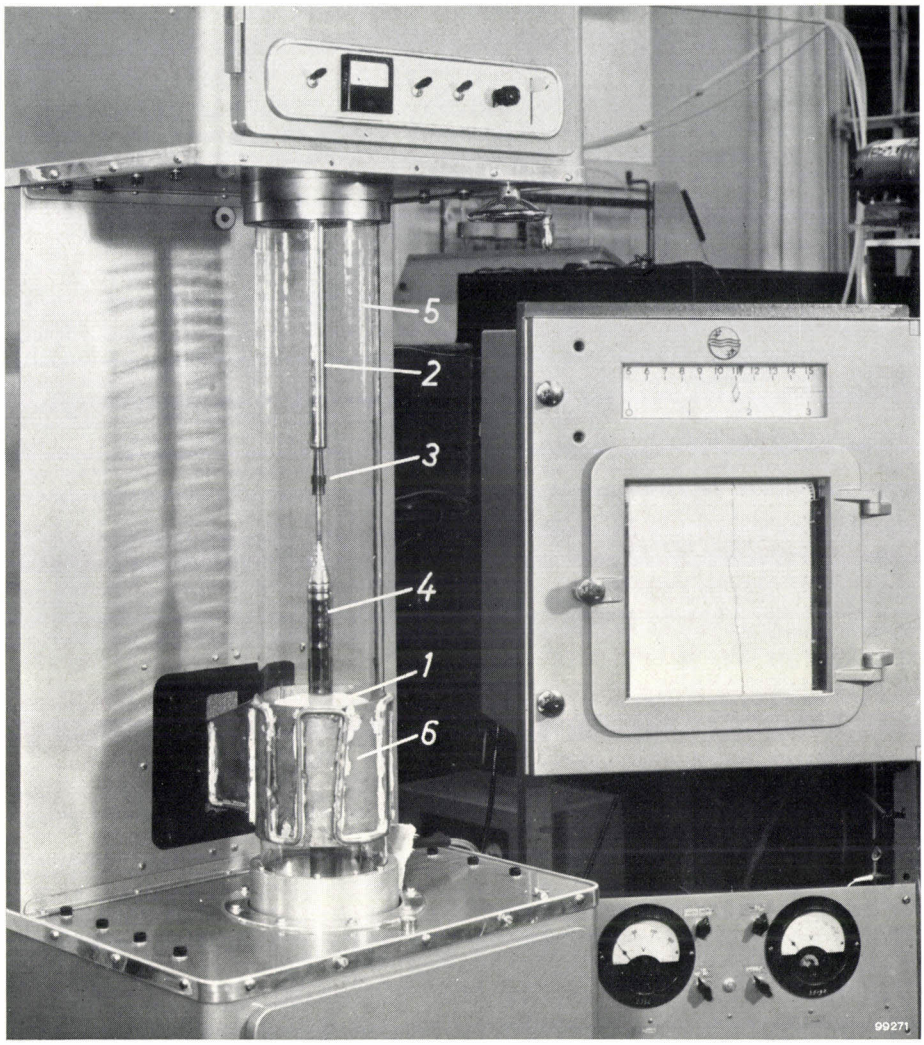


Fig. 5. Apparatus for pulling a single crystal from the melt (Czochralski method). *1* crucible containing molten germanium. *2* rotating shaft. *3* seed-crystal holder. *4* pulled single crystal. Components *1* to *4* are contained inside a glass envelope *5* through which an inert gas flows. The high-frequency heating is provided by a coil *6* in the form of a water-cooled plate encircling the glass envelope at the position of the crucible. The automatic potentiometer (right) holds the crucible at the desired temperature by automatic control of the heating supplied by the high-frequency generator; it also records the actual temperature.

of constant diameter grows on the seed crystal (*fig. 5* and *6*).

The situation at the solid-liquid interface in this case is similar to that in the solidification process described above, in which the liquid was stirred except in a layer of thickness δ. The magnitude of δ is here determined by the viscosity of the liquid and the speed of rotation of the shaft. The value assumed by the segregation coefficient k is therefore determined here by these quantities as well as by the speed at which the crystal is pulled and by the diffusion constant.

It may be concluded from the above considerations that, here too, C_S is nowhere in the single-crystal rod independent of position. Since the pulled crystal is the final product of the preparation process, this is a very undesirable state of affairs.

The equations given earlier [1]) for the solidification process were in fact first derived for the method of crystal-pulling just discussed. It should be noted here that the solution (5) is in fact an approximation, since it is an over-simplification to assume that there exists in the liquid a layer of thickness δ in which diffusion is alone responsible for impurity transport and which is sharply divided from the remainder. If the problem be posed exactly (that is to say for the pulling method with rotating crystal), it can only be solved numerically. For the point $x = 0$ the solution can be written in the form of an integral equation, but it is impossible to derive from this a formula for k.

The validity of the approximate treatment given above is corroborated by an experiment based on the following considerations. Since a single crystal is required, the temperature $T(x)$ in the liquid must everywhere be higher than the equilibrium temperature T_E which can be read for a given concentration from the phase diagram. If this condition is not satisfied, multiple nucleation will occur. Assuming that there is no supercooling — at the interface the temperature is necessarily equal to T_E — the requirement $T(x) > T_E$ in the immediate proximity of the interface ($x = 0$) can be replaced by:

$$\left(\frac{dT}{dx}\right)_{x=0} > \left(\frac{dT_E}{dx}\right)_{x=0}. \quad \ldots\ldots \quad (7)$$

The right-hand side of (7) can be written:

$$\left(\frac{dT_E}{dx}\right)_{x=0} = \left(\frac{dT_E}{dC_L}\right)_{C_L = C(0)} \times \left(\frac{dC_L}{dx}\right)_{x=0}. \quad \ldots \quad (8)$$

Fig. 6. Pulling a single crystal from the melt. The operation is almost complete. The photograph shows clearly the shape of the crystal obtained.

From (3) it follows that

$$\left(\frac{dC_L}{dx}\right)_{x=0} = -\frac{f\{C_L(0) - C_S(0)\}}{D} = -\frac{f}{D} C_L(0)(1 - k_0), \quad (9)$$

so that

$$\left(\frac{dT_E}{dx}\right)_{x=0} = -\left(\frac{dT_E}{dC_L}\right)_{C_L = C(0)} \times \frac{f}{D} C_L(0)(1 - k_0). \quad (10)$$

Substituting in this expression the values of dT_E/dC_L, D and k_0 for a given impurity and the relevant value of f, and putting dT_E/dx equal to the value found for dT/dx, we can calculate from (10) the maximum value that $C_L(0)$ may have under the given circumstances without the occurrence of nucleation.

In the case of germanium containing antimony as impurity, we have:

$$\frac{dT_E}{dC_L} = 400 \text{ °C/mol},$$

$$k_0 = 0.003,$$

$$D = 5 \times 10^{-5} \text{ cm}^2/\text{sec}.$$

Suppose $f = 1.8$ mm per minute or 3×10^{-3} cm/sec. Assuming that dT/dx in the solid is about 150 °C/cm at the solid-liquid interface, we can draw up an equation between the rates of heat transport to and from the interface ($x = 0$); we thus calculate that dT/dx in the liquid is approximately 28 °C/cm (taking the latent heat of fusion as 98 cal/gram and the thermal conductivity coefficient of solid and liquid Ge as 0.034 and 0.12 cal/sec.cm.°C respectively).

For $C_L(0)$ this gives a value of 1.4×10^{-3} mole/cm³, and hence for $C_S(0)$ a value of 4.2×10^{-6} mole/cm³, that is to say 2.6×10^{18} atoms of antimony per cm³. Calculation shows that the resistivity of the germanium with this value of C_S should be 0.004 Ωcm. Measured on a piece of germanium doped with Sb, the resistivity at the boundary between the single-crystal and polycrystalline region, i.e. at the position where T is still just above the value T_E, is found to be 0.003 Ωcm. Since the values taken for dT/dx and for the thermal conductivity are not known exactly, and since some supercooling occurs before nucleation begins [3]), the agreement is highly satisfactory.

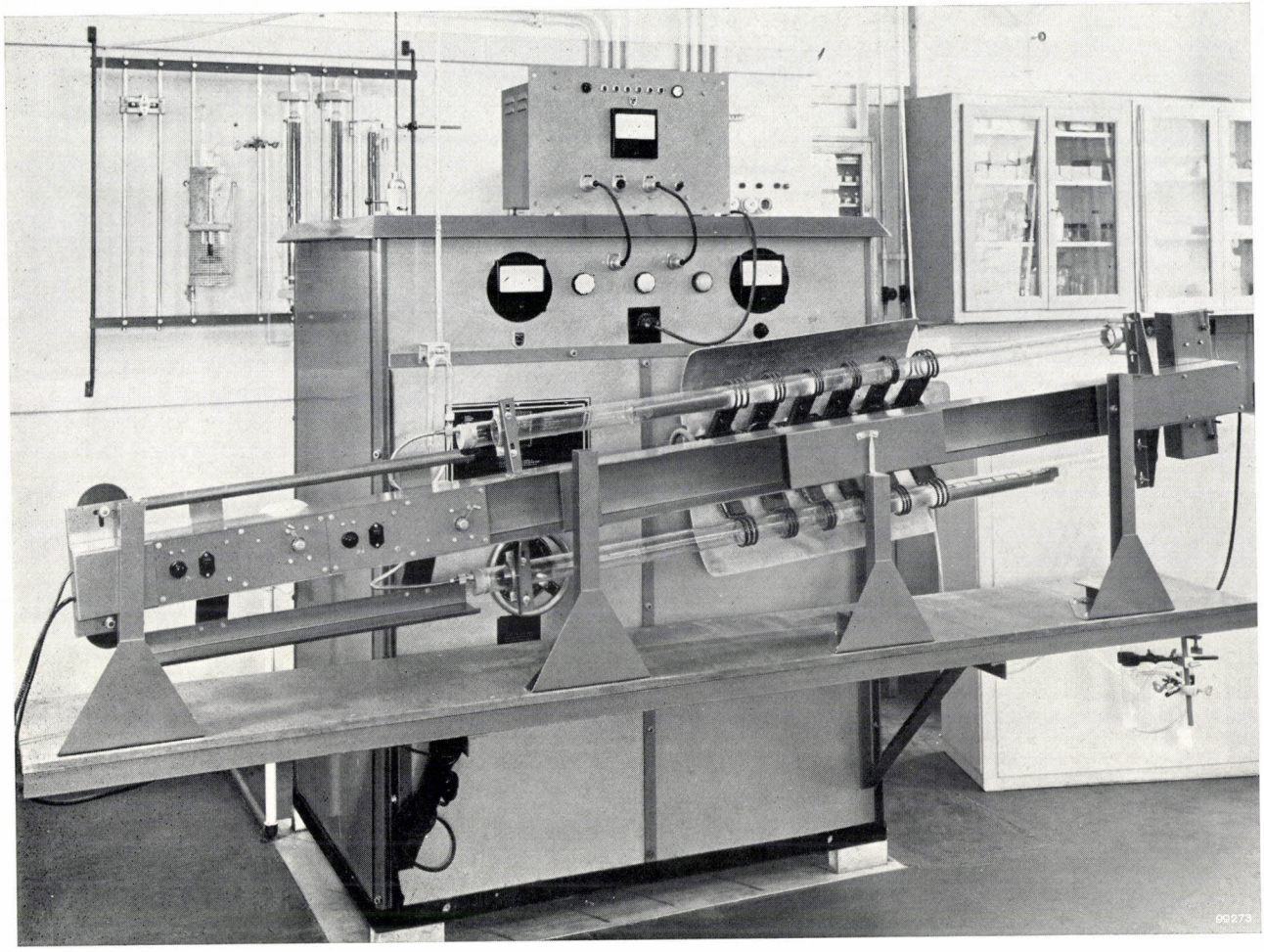
Modern doping and purification methods

The modern methods of doping germanium and silicon are all variants of the crystal-pulling method or of zone melting; purification is always effected by a zone-melting process. In the latter the ingot is locally heated so as to melt only a short zone of its length. The heating is usually inductive with the aid of a high-frequency generator, the molten zone being made to traverse the ingot slowly from one end to the other.

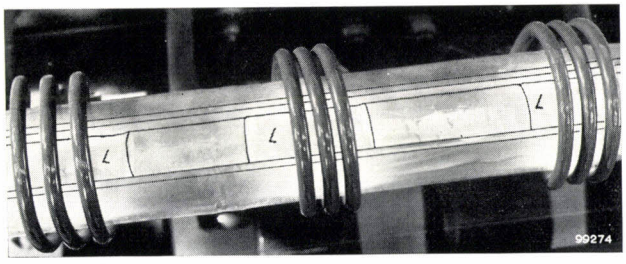
If we compare this process with that of causing one end of an entirely molten ingot to solidify, we notice that the amount of impurity segregated from the solidified part — we again assume $k_0 \ll 1$ — is taken up in a smaller volume of liquid, so that the increase in C_L is more pronounced and the purification less effective. Whereas in normal solidification C_L does not reach twice its initial value before half the ingot is solidified, in this case it reaches this value after only slightly more than two zone lengths.

The fact that zone melting is nevertheless preferable as a method of purification is due to the following circumstances. In the first place the process can be repeated indefinitely without it being necessary to cut off the part last to solidify, in which C_S is, of course, very large. Apart from the time saved, this is an advantage inasmuch as there is no risk of contaminating the ingot by any mechanical or manual handling. Further, a second and third zone can start to traverse the ingot before the first has reached the end. *Fig. 7* shows photographs of a zone-refining equipment for germanium in which the ingot passes through no fewer than six heaters, so that in a single run six molten zones pass through the ingot. This obviously gives an

[3]) B. Chalmers, Proc. Int. Conf. on Crystal Growth, Cooperstown (N.Y.) 1958, p. 297, Wiley, New York.



a



b

Fig. 7. *a*) Apparatus for purifying germanium by multiple zone melting. The graphite boats containing the germanium are enclosed in tubes through which dry purified hydrogen flows. During the zone-melting process the boats move slowly from right to left; in each run they pass six high-frequency heating coils situated at short distances apart. They are pulled forward by means of a thin metal wire. The tubes are on a slight slope to prevent the solidified charge from assuming a wedge shape due to the differing specific volumes of liquid and solid. At the end of a run the boats can be rapidly returned by mechanical means to their starting point in order to start a second run, and so on. The equipment on the wall rack is used for purifying the hydrogen.
b) Detail of the apparatus. The molten zones (L), at the position of the heating coils, are 4 to 5 cm long.

enormous saving of time. Nor is it necessary to handle the charge if it is to be passed through for a further run.

Since molten silicon is chemically highly reactive, melting it in a crucible gives rise to problems. Nevertheless it can be processed successfully by a special method of zone melting known as the floating-zone technique. In this method the charge is clamped at its ends in a vertical position. Provided it is not too long, the zone is held together by its own surface tension (*fig. 8*). Here, too, the growing crystal is made to rotate. A single-crystal product is obtained provided the end from which the molten zone begins is a single-crystal seed in contact with the charge proper. The floating-zone technique is the best method of making single-crystal silicon.

It may be inferred from what has been said that, in the forms discussed above, zone melting for the purposes of doping is open to the same objections as the methods outlined in the introduction. The concentration C_S is not constant over any appreciable length of the product.

An improvement is possible, however, by modifying the method. Where k is very small (<0.01)

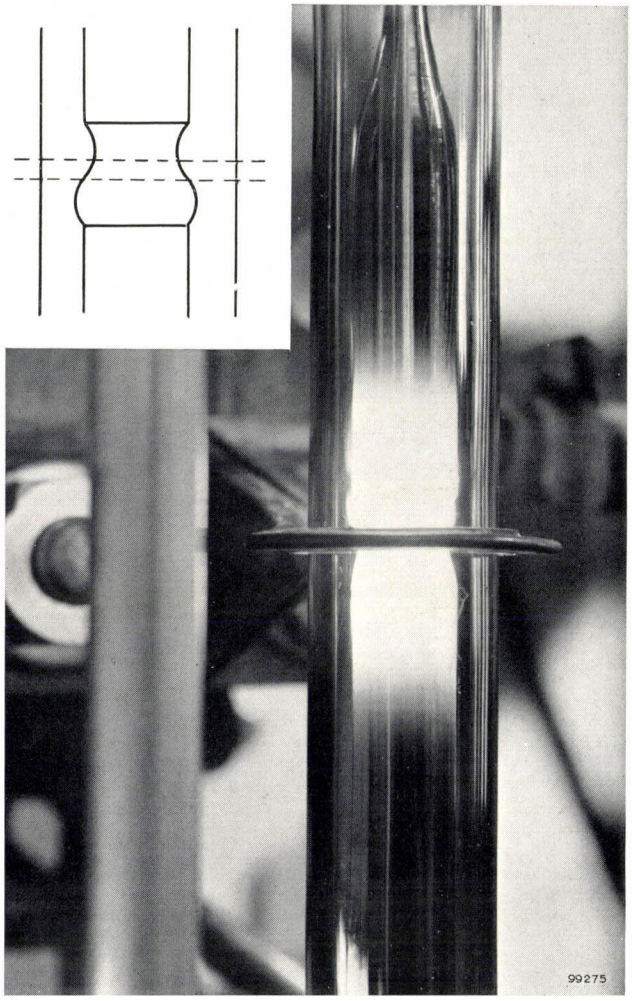


Fig. 8. Purification of silicon by the floating-zone technique. The charge is here set up vertically, clamped at the ends, in the middle of a relatively wide tube through which an inert gas flows. The (single) zone is held together by the surface tension of the molten material (see inset).

good results can be obtained, for example, by arranging for the first zone of the ingot to consist of impure material (concentration C_i) and the remainder of pure material. Except of course in the last zone, one then obtains approximately the concentration $k_0 C_i$ over the entire ingot. If k_0 is not very small, improvement is possible by causing a zone to pass to and fro many times, or by making use of the fact that C_S also depends on the length of the zone; in the latter case one aims at keeping C_S constant by changing the zone length during the course of the process. Use can also be made of the fact that k depends on the speed at which the zone moves through the ingot [4]). The first process is, of course, time-consuming, and all three are subject to practical difficulties.

The best doping methods are essentially those in which a highly-purified material is used and just as much of the required impurity is continuously supplied from outside to the molten zone as is withdrawn from the trailing solid-liquid interface. A product is then obtained in which C_S is entirely uniform throughout, irrespective of the value of k_0. We shall now discuss two processes developed in this laboratory which are based on this principle and have proved their usefulness. The first concerns the preparation of silicon doped with phosphorus [5]) — in this case $k_0 = 0.35$ and the method with the enriched first zone is therefore unsuitable — and the second the preparation of doped germanium [6]). The latter process can in principle be used for introducing virtually any impurity.

Phosphorus-doping of silicon by zone melting with continuous phosphorus feed

In the uniform vapour-phase doping of silicon with phosphorus the continuous phosphorus feed is effected by adding a trace of phosphine (PH_3) to the inert gas flowing around the ingot (the zone melting is done here of course by the floating-zone technique). The passage of the hot zone causes the PH_3 to decompose, and a certain fraction of the liberated P atoms is dissolved in the molten silicon. In the steady state the zone takes up as many P atoms per unit time as leave it via the trailing solid-liquid interface. The initial ingot of silicon must be well purified beforehand. The volume of the zone must remain constant during the process.

At the beginning of the process the ingot is surrounded only by a stream of inert gas. When a molten zone has formed, PH_3 is added at a constant rate to the stream of gas. As soon as the P concentration in the zone has attained the required value — we shall see presently how this is ascertained — the ingot (that is to say the zone, seen from the ingot) is set in motion.

Fig. 9 shows a plot of resistivity versus position for various single-crystal silicon rods produced in this way. It can be seen that over a very considerable length of each rod the resistivity deviates by no more than about 10% from the average value. The

[4]) The various methods of zone melting are discussed in W. G. Pfann, Zone melting, Wiley, New York 1958.

[5]) J. Goorissen and A. H. J. G. van Run, Gas-phase doping of silicon, Proc. Instn. Electr. Engrs. 1959, in press.

[6]) J. Goorissen and F. Karstensen, Das Ziehen von Germanium-Einkristallen aus dem „schwimmenden Tiegel", Z. Metallkunde **50**, 46-50, 1959 (No. 1), and J. Goorissen, F. Karstensen and B. Okkerse, Growing single crystals with constant resistivity by floating-crucible technique, published in Solid state physics in electronics and telecommunications, Proc. int. Conf. Brussels, June 2-7, 1958, edited by M. Désirant and J. L. Michiels, Vol. 1, pp. 23-27, Academic Press, London 1960.

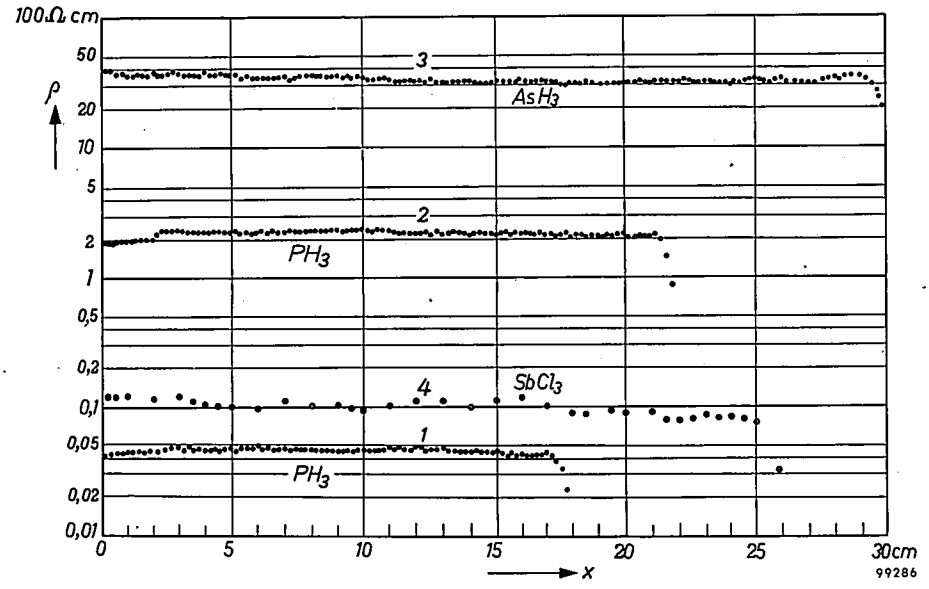


Fig. 9. Series *1* and *2*: resistivity ϱ as a function of position (x), measured on silicon rods doped with phosphorus; the impurity was added in the floating-zone treatment by means of a constant stream of PH_3. Except in the last zone, ϱ differs nowhere by more than 10% from the average value. Since for silicon containing phosphorus k_0 is 0.35, this degree of constancy of ϱ is very difficult to achieve by other methods. Series *3* and *4*: as above, but for silicon doped in the same way with arsenic (via AsH_3) and antimony (via $SbCl_3$), respectively.

same figure shows some provisional results of experiments using the same method for the doping of silicon with arsenic (via AsH_3) and with antimony (via $SbCl_3$).

An investigation into the principles governing the uptake of phosphorus in the molten zone has shown that, other conditions being equal (i.e. geometry, rate of zone travel and speed of rotation), the amount taken up per second in the growing rod depends solely on the amount of PH_3 flowing to the zone per second. If this amount is increased by, say, a factor of 2, the conductivity of the rod increases by the same factor; see *fig. 10*. It is immaterial in this connection whether the increase is brought about by doubling the PH_3 concentration with the gas flow remaining constant, or by increasing the gas flow with the concentration remaining constant. Finally, if we increase the speed of the zone, other conditions remaining equal, the phosphorus concentration in the rod decreases by the same factor. This means that the rate at which phosphorus is absorbed in the zone does not depend upon the rate at which the zone travels; the rate of absorption is also evidently independent of the phosphorus concentration in the zone at any instant. It may therefore be assumed that the rate of absorption during the initial period, in which the molten zone is not yet moving, is the same as when the zone is in motion. It is thus possible to calculate the time that must elapse, after introducing the PH_3 stream, before the zone can be started moving.

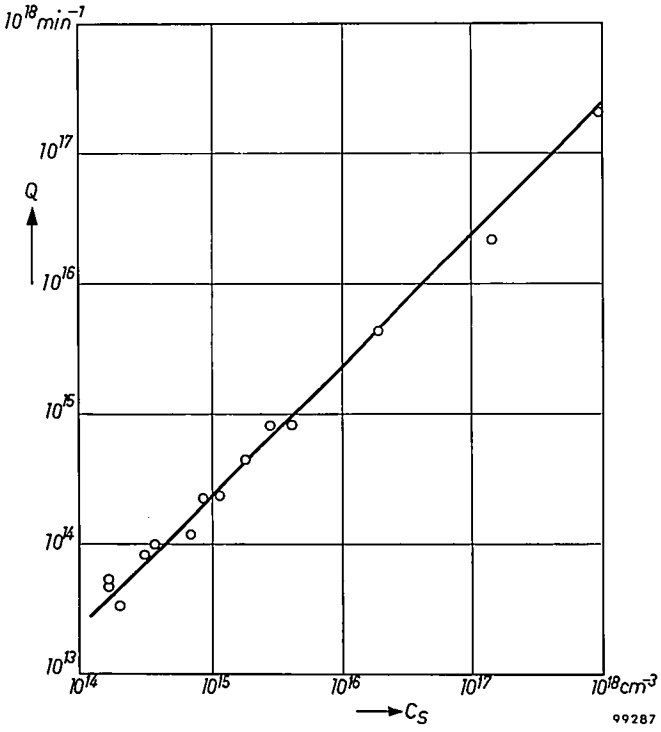


Fig. 10. Relation between the logarithm of the phosphorus concentration C_S in a silicon rod and that of the PH_3 stream Q to the zone. Over a very large region C_S and Q are related by a straight line of slope unity; the phosphorus concentration is therefore directly proportional to the quantity of PH_3 flowing per unit time. The units employed are P atoms per cm^3 for C_S, and PH_3 molecules per minute for Q.

A floating-crucible technique for the preparation of homogeneously doped germanium crystals

The floating-crucible technique [6]) for the preparation of doped single-crystal germanium with a homogeneous distribution of impurity is a variant of the Czochralski crystal-pulling method. Although it differs considerably from the floating-zone method just described, the basic idea is the same: in this case too a zone of constant volume is present to which as much impurity is added per second as leaves to enter the growing crystal from the solid-liquid interface. Again, the method is practicable whatever the value of k_0.

The method is illustrated schematically in *fig. 11*. The molten material *1* from which the single crystal *2* is pulled, is contained in a small graphite crucible *3* which floats on the melt *4* contained in a second crucible *5* mounted on legs. The contents of both crucibles are in communication via a capillary *6*. A stem underneath the floating crucible passes through a hole in the bottom of the large crucible and carries a weight *7*. The surface tension of germanium is sufficient to prevent the liquid from flowing away through the gap between the stem and the bottom of the large crucible.

The procedure is to fill the large crucible with pieces of pure germanium and to melt them, whilst the small crucible lies at the bottom. It remains there after the germanium is melted. The desired impurity is then added to the melt until its concentration is equal to the value C_S required in the product. With the spindle *8* the small crucible is now pushed up until it partly projects above the surface of the melt. The crucible then remains floating because the hydrostatic upward (buoyancy) force is now supplemented by an upward force due to the surface tension. Owing to the slope of the side walls of the small crucible, this capillary force varies with the depth of submergence. As a result vertical movements, due for example to external disturbances, remain small in amplitude.

Next, the rotating seed crystal is lowered on to the surface and the process of pulling a single crystal begins. This gives rise to a flow of liquid in the capillary. Thereupon an amount of the desired impurity is added to the germanium in the *small* crucible until the concentration there rises to C_S/k. Since the volume of the liquid in this crucible does not depend on the level of the liquid in the large crucible — we shall explain the reason presently — from that moment a stationary state is set up, provided the capillary is so narrow that the number of impurity particles leaving the floating crucible per second by diffusion is negligible compared with the number flowing in with the germanium. (It can be shown that this is the case if the rate of flow of the germanium in the capillary is high compared with the ratio of the diffusion constant to the length of the capillary.) Apart from a short region at the beginning and end, the single-crystal rod thus obtained possesses over its whole length a constant impurity concentration equal to the desired value C_S.

The constancy of the quantity of material in the small crucible is due to the fact that the stem carrying the weight *7*, and also the part of the crucible vertically above it, make no contribution to the hydrostatic upward force. This remains constant even though the liquid level drops in the large crucible. The submergence of the floating crucible does not therefore change as the operation proceeds; the volume of germanium therein remains constant.

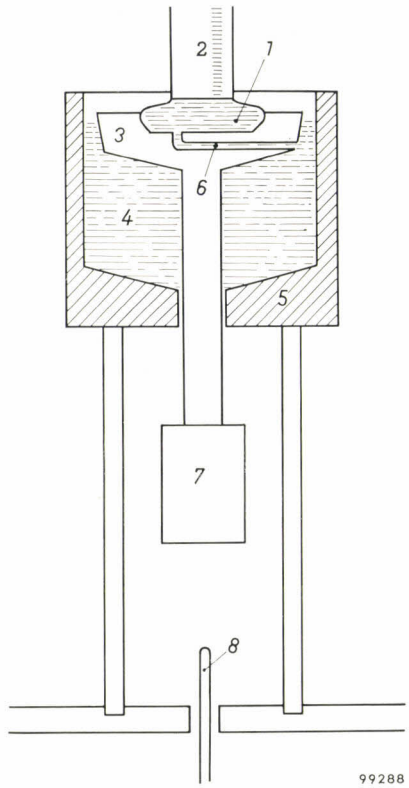


Fig. 11. Schematic cross-section of an arrangement for making single crystals of germanium with constant impurity concentration (floating-crucible technique). *1* molten germanium from which the crystal *2* is pulled and which is contained in the crucible *3*. This floats on a much larger quantity of germanium *4* with which the crucible *5* is filled. *6* connecting capillary. *7* weight. *8* spindle for pushing up the crucible *3* after the germanium in *5* is melted.

As might be inferred from the foregoing, the weight *7* is so chosen that the total weight of the small crucible, the stem and *7* is greater than the hydrostatic upward force. Consequently the small crucible not only remains at the bottom when the large one is being filled — this is necessary in order

to fill the capillary, though that could of course also be done without the weight 7 — but also the submergence when floating is sufficient. In the absence of 7 the small crucible would not sink deeply enough, and therefore it would lose all its liquid.

From the point of view of apparatus and procedure the method described here is closely related to the ordinary method of crystal pulling. Having regard, however, to the various volumes involved (those of the pulled crystal, the contents of the floating crucible and the contents of the large crucible) and considering the concentrations, it is evident that it can in fact be regarded as an unusual form of zone melting. The contents of the small crucible play the part here of an artificially enriched molten zone, and the large crucible can be thought of as the initial charge. The latter, however, already possesses the desired concentration (which is moreover independent of position), so that as the operation progresses the "zone" does not change in concentration.

Fig. 12 shows the result of resistivity measurements as a function of position for various germanium rods made by this method. Here, too, the resistivity in a large region differs by less than 10% from the average value.

Finally, a remark about the effectiveness of the two new methods discussed above. By this is meant the percentage of the total rod length in which the extreme resistivity values do not differ by more than, say, 20%.

Assuming that the concentration necessarily differs from the desired value only in the last zone, but is otherwise everywhere constant, the effectiveness of the two new processes is theoretically equal

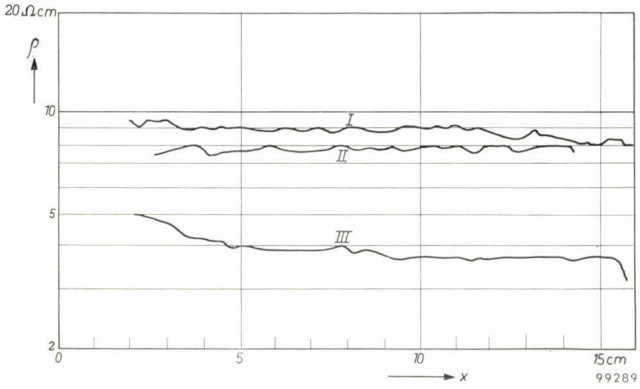


Fig. 12. Resistivity ϱ as a function of position (x), measured on various germanium crystals made by the floating-crucible technique. The impurities used were: curve I — indium ($k_0 = 0.0014$); curve II — antimony ($k_0 = 0.007$); curve III — phosphorus ($k_0 = 0.12$). Over most of the length of each crystal the resistivity ϱ is constant to within 10%.

to $100(1-V_z/V)$ %, where V_z is the volume of the zone (or of the contents of the floating crucible) and V is the total volume. In the floating-zone method, where silicon is continuously doped with phosphorus, this theoretical effectiveness, which is 90% for a rod ten zones long, can in fact be achieved. It is not quite realizable, however, with

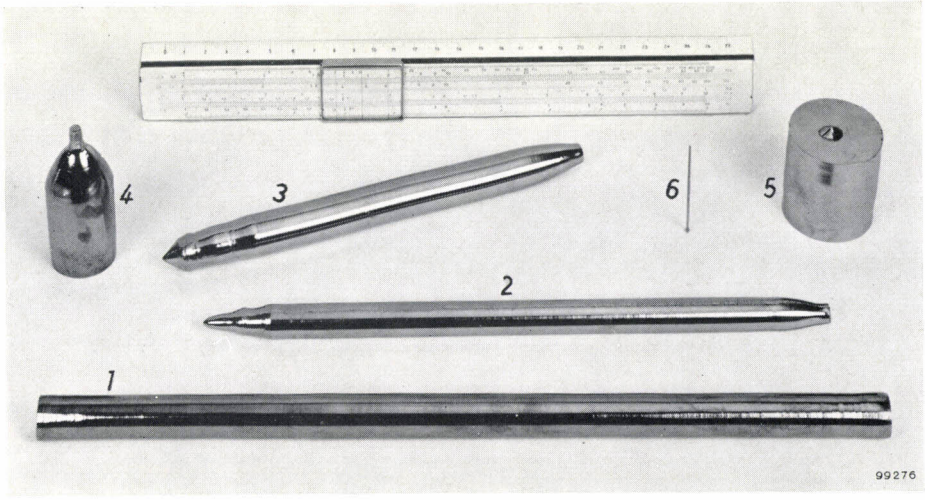


Fig. 13. Single crystals of various dimensions, made by the methods discussed in this article. *1* silicon crystal with constant phosphorus concentration (floating-zone melting with PH_3 doping), *2* and *3* germanium (floating-crucible method), *4* top end of a thick germanium crystal (floating crucible), *5* bottom end of a very thick germanium crystal (normal pulling method), *6* very thin germanium crystal (idem). From *1* and *2* it can be seen that the crystals do not have a circular cross-section. This is due to the fact that the rate of growth is not the same in all crystallographic directions.

the floating-crucible method. In the first place the volume of the liquid in the small crucible depends on the diameter of the rod (owing to the effect of the surface tension) and is therefore not constant at the beginning while the diameter of the crystal is still growing. In the second place, not all the germanium in the large crucible can be used. The effectiveness obtainable in practice with our apparatus is found to be about 75%; the theoretical efficiency is 85%.

Just as high an effectiveness can also be obtained with the earlier-mentioned zone-melting method using an enriched first zone, but only if k_0 is smaller than about 0.01; with increasing k_0 the efficiency drops sharply.

The effectiveness of the ordinary crystal-pulling method — we refer to germanium of course, silicon being handled without a crucible — is poor in the majority of cases. By varying the pulling rate, it can be raised to a theoretical value of 73% [7]) in the most favourable case (viz. with antimony, where k varies most strongly with the pulling rate). However, the complications involved are considerable.

Fig. 13 shows a photograph of a silicon rod and various germanium rods. All are single crystals; the first was obtained by the floating-zone technique; all but two of the germanium rods were produced by the floating-crucible technique.

[7]) Discussed at length in the first of the articles quoted under [6]).

Summary. At the solid-liquid interface of a substance containing a very small amount of an impurity, the concentration of this impurity in the solid phase generally differs from that in the liquid phase. The ratio k_0 between both concentrations is called the distribution coefficient. As the solidification progresses, the segregation constant k, i.e. the ratio between the concentration in the solidified part and that in the liquid (away from the solid-liquid interface), lies between k_0 and 1. If the liquid is vigorously stirred, then $k \approx k_0$; if there is no convection, then $k = 1$. The segregation process can be used for purification ($k < 1$), but it makes it difficult to obtain a homogeneous product. Nowadays, silicon and germanium are purified by the method of multiple-pass zone melting, whereby fairly short molten zones are passed from one end of an ingot to the other. In germanium, which can be melted in a crucible, the successive zones are passed through at short distance apart, and this represents a considerable saving of time. Silicon in the molten state is too reactive to be melted in a crucible. In this case one zone is passed through a vertical charge, clamped at either end, the zone being held intact by its own surface tension (floating-zone technique). The segregation phenomenon is troublesome if the material is to be doped. In general, a homogeneous product is not obtained (unless k_0 is very small and an enriched first zone is used). Two new methods have been evolved in the Philips laboratories. They are based on the idea that, irrespective of the value of k_0, a homogeneous product is obtained if as much impurity is added from outside to the zone per second as leaves it to enter the growing crystal via the solid-liquid interface. A homogeneous rod of single-crystal silicon, doped with phosphorus (k_0 is 0.35 for phosphorus in silicon), can be produced by adding a trace of PH_3 to the inert gas flowing around the charge. In the steady state the molten zone takes up as much phosphorus from the gas — PH_3 dissociates at high temperatures — as is expelled from it via the trailing solid-liquid interface. Homogeneously doped single-crystal germanium (impurity concentration C_S) can be made by the floating-crucible technique, which is a variant of the ordinary method of crystal pulling. The crystal is pulled from the contents of a small crucible, concentration C_S/k, which communicates via a capillary with the contents of a larger crucible of concentration C_S. As the operation progresses, the contents of the small crucible change neither in volume nor in concentration. The effectiveness of both methods as regards the homogeneity of the product is high and approaches close to the theoretical value, viz. $100(1-V_z/V)$ %, where V_z is the volume of the zone (or of the melt, in the floating crucible) and V is the total volume.

AN AUTOMATIC DEW-POINT HYGROMETER USING PELTIER COOLING

by P. GERTHSEN *), J. A. A. GILSING and M. van TOL.

537.322.15:536.423.45:662.613.54

Progress in the field of semiconductors has in recent years opened up more and more new applications. The article below describes an automatic dew-point hygrometer in which the necessary cooling is produced by the Peltier effect of a certain combination of semiconductors.

The most accurate method of measuring the humidity of air and other gases still remains the determination of the dew point. Instruments based on other commonly used methods are either dependent on material properties and must therefore be regularly calibrated (hair-hygrometers, diffusion method), or they are liable to get dirty, without the resultant errors becoming perceptible (electrolytic methods, wet and dry bulb). The other methods proposed, such as weighing before and after drying, measuring the speed of sound or the dielectric constant are scarcely suitable for industrial application.

A complication of the dew-point method is the cooling required in order to reach the dew point. For this reason the dew-point method has hitherto been used only in laboratories, and has not been adapted to industrial use.

As a result of intensive solid-state research in recent years, semiconductors are now available which make electrical cooling possible by very simple means, based on the use of the Peltier effect.

This effect (one of the group of thermoelectric effects) occurs when an electric current is passed through a circuit of two conductors of dissimilar material connected in series. Heat is generated at the one junction, whilst heat is absorbed at the other. The amounts of heat absorbed or developed per unit time are proportional to the current. The proportionality factor is the Peltier coefficient P of the particular combination of conducting materials. If the situation is reversed, i.e. if the circuit is opened and the junctions are brought to different temperatures, there appears across the terminals of the thermocouple so produced a potential difference that depends on the temperature of each of the junctions. The latter phenomenon, called the Seebeck effect or more generally "the" thermoelectric effect, is commonly used for temperature measurement [1]).

*) Zentrallaboratorium Allgemeine Deutsche Philips Industrie G.m.b.H., Aachen Laboratory.

[1]) For an introduction to the various thermoelectric phenomena, and a recent review of developments, see A. F. Ioffe, Semiconductor thermoelements and thermoelectric cooling, Infosearch, London 1957.

The apparatus for determining the dew point is represented schematically in *fig. 1*. Blocks *1* and *2*, made of different semiconducting materials, are joined at positions *a* by a silver plate *3*, about 5 mm thick and wide and about 15 mm long. At positions *b* they are connected to fairly large and thick copper plates which act as "cooling fins" and also conduct

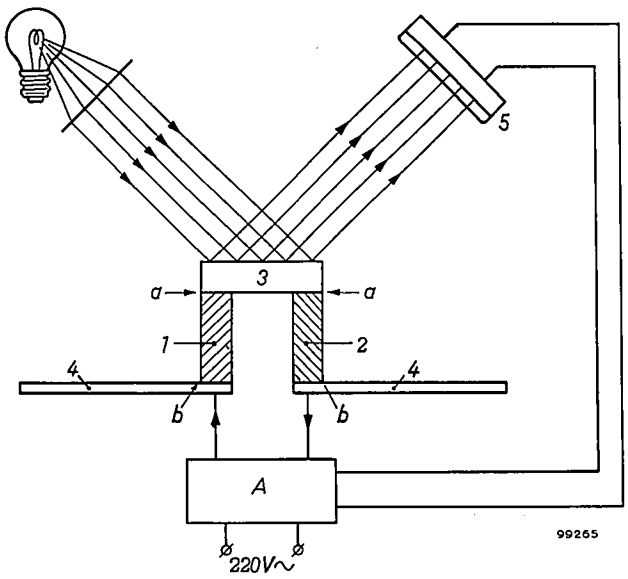


Fig. 1. Schematic representation of apparatus for automatic dew-point determination. Components *1*, *2*, *3* and *4* together constitute the whole Peltier element. *1* and *2* are two different semiconductors. *3* silver plate with reflecting surface. *4* copper plates acting as cooling fins for the warm junctions; they also serve as electrical terminals. When a current flows, heat is absorbed at the junctions *a* and developed at the junctions *b*. As the mirror mists over, less light falls on the photoresistor *5*, and the circuit *A* then feeds a smaller current through the Peltier element. In the state of equilibrium the temperature of *3* is equal to the dew point.

the current to the blocks. The junctions *b* are thus always at about room temperature. The upper surface of the silver plate is polished so as to produce a plane mirror. A parallel beam of light, obtained from an electric bulb and a lens, is directed on to this mirror at an angle of approximately 45°. The reflected beam impinges on a small photoresistor of cadmium sulphide [2]). When the apparatus is

[2]) See N. A. de Gier, W. van Gool and J. G. van Santen, Photoresistors made of compressed and sintered cadmium sulphide, Philips tech. Rev. **20**, 277-287, 1958/59 (No. 10). The resistor used here is of the type shown on the right in fig. 11 of that article.

switched on, so that current flows through the blocks and the reflector in the right direction, heat is developed at the junctions *b* and withdrawn from the junctions *a*. (Just as much heat is withdrawn from the junctions *a* as if the blocks *1* and *2* were directly connected with each other; the same applies to *b*.) As soon as the temperature of the silver plate drops below the dew point, the surface of the strip mists over and ceases to reflect specularly. The incident light is now reflected diffusely and less light reaches the photoresistor. As a result, the current flowing through this resistor also decreases. By means of an appropriate circuit, represented in the figure by *A*, the photocurrent controls the current flowing through the Peltier element in such a way that the cooling diminishes when the reflector mists over. The entire system thus constitutes a closed control loop. The temperature of the silver plate then rises once more above the dew point, the mist disappears and the process repeats itself. Effective damping of the control loop is ensured by suitably selecting the time constants, so that the equilibrium temperature is reached after only one or two oscillations; this temperature is the dew point.

It should be added that the diffuseness of reflection of the misted reflector depends on the "thickness" of the layer of condensate. This thickness, which we can express as the quantity of water per cm^2 of reflector surface, is proportional to the difference between the temperature of the reflector and the dew point and proportional to the time. After the temperature has dropped below the dew point, the nett cooling decreases but does not become negative *immediately:* the control is not an on-off but a proportional control, and the "integration" of dew on the mirror gives the system a measure of integral action [3]).

The fact that the equilibrium temperature coincides with the dew point and is not below it (it could not possibly be above it, because the reflector is then not misted and full cooling takes place) can be understood as follows. If the temperature of the reflector, and hence of the water condensed on its surface, is below the dew point, the mist will become increasingly dense and the cooling will decrease. Only when this temperature rises to the dew point can the mist achieve a steady state and the cooling remain constant.

[3]) See for example H. J. Roosdorp, On the regulation of industrial processes, Philips tech. Rev. **12**, 221-227, 1950/51. A distinctive feature of integral action in control systems is that the difference between the desired value and the value which the controlled quantity actually assumes in the state of equilibrium — the offset — is necessarily equal to zero.

The error in the dew-point determination with a hygrometer of this type can easily be made less than $\frac{1}{2}$ °C.

A working form of the circuit *A* (fig. 1) is illustrated in *fig. 2*. Since the resistance of the Peltier element *P* is low but the current through it fairly high (as much as 7 A), the simplest method of obtaining this current is to use an A.C. mains supply

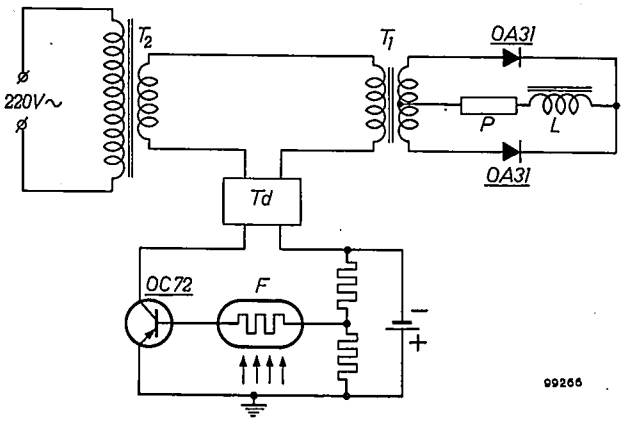


Fig. 2. Simplified diagram of the circuit. *P* Peltier element. *L* choke. T_1 and T_2 transformers. *Td* transductor (saturable reactor). *F* photoresistor, type B 873103. There are also two OA 31 germanium diodes and a type OC 72 transistor.

transformed down to a low voltage and then rectified. The primary of a transformer T_1 is connected via a transductor (saturable reactor) *Td* to the secondary of the mains transformer T_2. The direct current is obtained by means of two germanium diodes, type OA 31, and is smoothed by a choke *L* (capacitors are not so suitable for this purpose). The control current flowing through the saturable reactor is varied by a photoresistor *F* (type B 873103) with the aid of a transistor type OC 72.

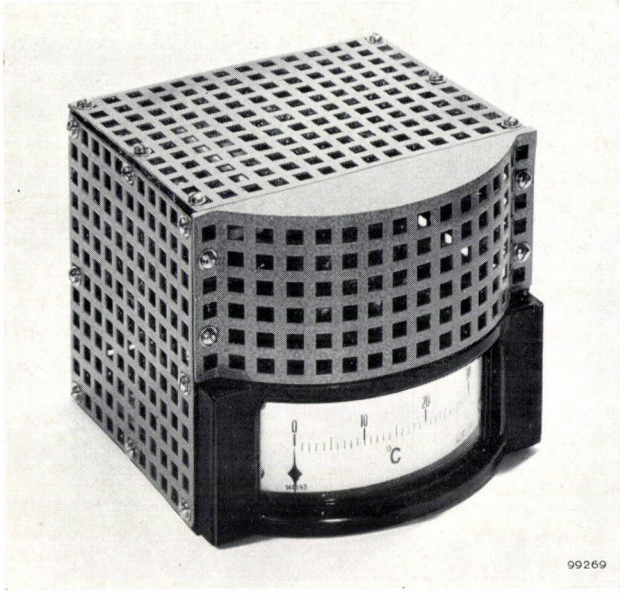
In the side of the silver plate are a number of 1 mm holes into which a thermocouple or thermistor is introduced for measuring the temperature.

A thermocouple made of the same materials as the Peltier element is particularly suitable. Such a thermocouple gives a very high e.m.f. (340 μV/°C) and has a low electrical resistance, so that the whole measuring circuit can easily be given a constant resistance; in this way it is possible to measure the current in the circuit instead of the e.m.f. The current may easily be as high as 10 mA.

A complication, however, is that a milliammeter having an internal resistance of less than 1 ohm is required, and moreover this resistance must not be temperature-dependent.

Fig. 3a shows a photograph of the complete apparatus, and fig. 3*b* the Peltier element with the copper plates acting as cooling fins. These plates stand vertically in the instrument, close to the rear

a



b

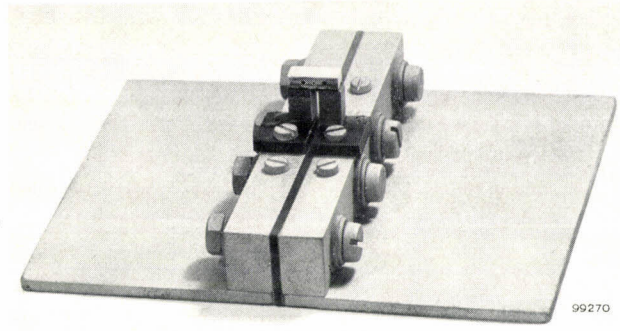


Fig. 3. *a*) Automatic dew-point hygrometer using Peltier cooling. *b*) Peltier element with copper plates acting as cooling fins. In the instrument these plates are vertically positioned close to the rear wall.

wall. The space above the millivoltmeter contains the lens system and the electrical circuit. The millivoltmeter is calibrated in °C.

We shall now consider at greater length the process of cooling by means of the Peltier effect. First, it should be noted that there is a limit to the temperature depression attainable with a Peltier element. True, the temperature depression produced is proportional to the current flowing through the element, but on the other hand ohmic heat is also generated, which increases in proportion to the *square* of the current. There is thus always a certain current at which the ohmic heating exactly compensates the Peltier cooling. Furthermore, there is also heat transfer from the warm to the cold junction via the Peltier elements themselves.

In order to draw up the heat balance from which we can calculate the maximum attainable reduction in the temperature of the reflector, it is necessary to make some simplifying assumptions:

a) The thermal conductivity and the electrical conductivity of both the materials constituting the element are equal (with the semiconductors employed this is in fact virtually true).

b) The cooling fins are so effective that the "warm" junction is at room temperature; the rest of the element is thus colder than the surroundings.

c) The quantity of heat absorbed from the surroundings is negligible.

d) The electrical and thermal resistance of the silver plate (*3* in fig. 1) and of the copper plates (*4* in fig. 1) are negligible.

e) Half of the ohmic heating is effectively dissipated at the junctions *a* and the other half at the junctions *b*.

f) All material constants are independent of temperature.

For the contributions to the power balance that refer to the cold junction (i.e. at the silver plate) we can now write the following simple formulae:

Peltier heat withdrawn: $W_1 = PI$.
Ohmic heat developed: $W_2 = \frac{1}{2} I^2R$.
Heat from the warm junction flowing per unit time through the Peltier elements: $W_3 = L\Delta T$.

(R is the electrical resistance of blocks *1* and *2* in series, and L is the thermal conductance of the blocks *1* and *2* in parallel.)

In the state of equilibrium W_1 must be equal to the sum of W_2 and W_3. Hence:

$$L\Delta T = PI - \tfrac{1}{2} I^2R. \quad . \quad . \quad . \quad . \quad . \quad (1)$$

By differentiating the right-hand side with respect to I and equating the result to zero, we find the current at which ΔT reaches its maximum value: $P - IR = 0$, whence $I_{max} = P/R$. From (1) we deduce that this value of ΔT is equal to:

$$\Delta T_{max} = \tfrac{1}{2} P^2/RL. \quad . \quad . \quad . \quad . \quad (2)$$

If, as is done in the dew-point hygrometer, the Peltier elements have the form of blocks of cross-section F and length l, it is a simple matter to express R and L in terms of the dimensions of the blocks. We find: $R = 2l\varrho/F$ and $L = 2F\lambda/l$, where ϱ is the resistivity and λ is the thermal conductivity of the materials used. In this case (2) becomes:

$$\Delta T_{max} = P^2/8\varrho\lambda. \quad . \quad . \quad . \quad . \quad (3)$$

It should be noted that the dimensions of the Peltier elements no longer occur in (3). Indeed, it can be shown that, whatever the form of the wires,

their dimensions are immaterial. If, for example, we try to reduce the ohmic heat by using short and thick blocks, we find that the "cold" gain is exactly offset by the increased flow of heat from the warm to the cold junction.

In cases where the above simplifying assumptions are not permissible, the foregoing is of course no longer true. The maximum temperature difference obtainable is then smaller than follows from (2) or (3). An important practical instance is where the Peltier element is required to supply a finite cooling *power*. Assumption *c*) is then no longer valid and a fourth term must be included in equation (1).

Combinations of substances for which the product $P^2/\varrho\lambda$ is relatively very large are to be found amongst the semiconducting compounds. A particularly large value of P can be obtained by combining an n-conductor with a p-conductor [4]). Of the two materials used in our case, one consists of 80% (by weight) Bi_2Te_3 + 20% Bi_2Se_3+0.02%AgI (n-conductor) and the other consists of 60% Sb_2Te_3 + 40% Bi_2Te_3 + 0.05% Ag. In our dew-point hygrometer they are used, as we have seen, in the form of blocks, 9 mm in length and of breadth and thickness 5 mm. At 3 A this element gives a temperature drop of about 30 °C. The maximum temperature drop is about 45 °C, which is reached when I is approximately 7 A [5]).

The above-described hygrometer using Peltier cooling is a variant of an automatic apparatus previously developed in our laboratory, which was based on the same principle (a change from specular to diffuse reflection caused by the misting of a reflector) but in which the cooling was effected by means of liquid air. It will perhaps be useful to give a brief description of this apparatus.

In this case the temperature of the silver reflector is automatically regulated by combining constant cooling with variable heating. The constant cooling is achieved in the following way. From a Dewar vessel filled with liquid air there protrudes one end of a bent rod of copper, about 10 mm in diameter. To the upper end is attached a disc of a material which is a poor thermal conductor (thermal resistance R_w), followed by a polished silver disc (*fig. 4*). The part of the rod projecting from the Dewar vessel is surrounded by thermal insulation. Because of its high thermal conductivity the entire rod assumes practically the temperature of liquid air (−194 °C). At the commonly occurring dew points (0-25 °C) the temperature difference between rod and reflector is thus always approximately the same, viz. about 200 °C: hence a practically constant heat flux passes from the silver disc to the copper rod. The thickness of the disc of insulating material is such that

$$200\ ^\circ\mathrm{C}/R_w = 2 \text{ watts}.$$

The (variable) heating is provided by means of a turn of resistance wire wound around the rim of the silver disc and electrically insulated from it by mica. Thermal insulation is provided on the outside. The heater wire is connected via an ordinary loudspeaker transformer to a power amplifier valve which is driven by a 50 c/s alternating voltage. By varying the amplitude of this alternating voltage the power supplied to the disc can be varied between 0 and 4 W. With constant cooling at 2 W the total power is thus variable between −2 and +2 W. (When the disc is cold, the heat it receives from the ambient air is of the order of some tenths of a watt, which is by comparison insignificant.)

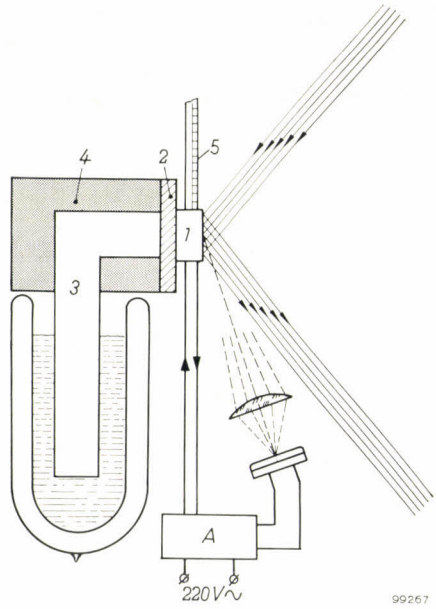


Fig. 4. Principle of dew-point hygrometer using liquid air. Constant cooling of the silver plate *1* is achieved by connecting it via a thermal resistance *2* to a copper rod *3*, the other end of which is immersed in a Dewar vessel containing liquid air; *4* thermal insulation, *5* mercury thermometer (may be replaced by a resistance thermometer or thermocouple). When the surface of *1* is clear, the beam of light bypasses the lens. When the surface mists over, some light is diffusely reflected on to the photocell connected to *A*; the latter then sends a current through the heater wire wound around *1*.

The automatic adjustment to the dew point is arranged in this case such that specularly reflected light from the clear surface of the silver disc just bypasses a lens in whose focal plane the light detector is situated. When the reflector becomes misty, the light is diffusely scattered. As long as the film of mist is still thin, however, most of the light is scattered in directions that make a fairly *small* angle with the direction of the reflected beam, so that quite a lot of light reaches the lens and hence the photocell or photoresistor. The relation between the photocurrent and the thickness of the film of mist is represented in *fig. 5*. At small thickness the photocurrent is proportional to the thickness of the film; it is in this region that the apparatus functions. As soon as light impinges on the photocell, the photocurrent gives rise to a voltage which drives the power valve referred to above, in such a way that the current through the heater wire (the anode current of the valve) increases as the mist thickens.

The characteristic difference between this instrument and the other is that in this case the current that regulates the temperature of the silver disc (the current through the heating

[4]) Concerning n and p conductors, see e.g. J. C. van Vessem, The theory and construction of germanium diodes, Philips tech. Rev. **16**, 213-224, 1954/55.

[5]) For comparison it may be noted that the combination of metallic conductors having the highest Peltier coefficient (Bi with Sb) can give a maximum temperature drop of just over 10 °C.

wire) must *increase* in conditions under which the current through the Peltier element should *decrease*, and vice versa. This is achieved by causing the specularly reflected light to bypass the light detector: the latter receives light only when the reflector is misted and thus reflects diffusely, the more so the thicker the film of mist. Here, then, the amount of light

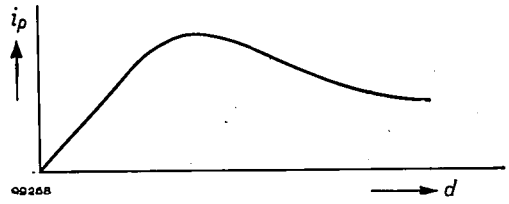


Fig. 5. Relation between the photocurrent i_p and the "thickness" d of the mist on the mirror. At small values of d the relation is linear.

incident on the detector varies from *zero* upwards, whereas in the apparatus using Peltier cooling it varies from a large to a smaller value. The former way of working may often be advantageous but is not used in the new hygrometer; the high sensitivity of the CdS photoresistor makes the other method no less satisfactory.

Finally it should be noted that the current through the resistance wire need not necessarily be a direct current, as required for Peltier cooling. It is thus possible to use a simple A.C. amplifier with transformer coupling to the output load; the latter can therefore be a short resistance wire through which a fairly large current flows. At the input of the amplifier, an alternating voltage is applied across the photocell.

In conclusion it may be said that the instrument which is the subject of this article is a nice example of the multifarious applications of semiconductors. They are used here (cf. fig. 2) not only in the Peltier element but also in the photoresistor, the transistors, the germanium diodes and finally in a resistance thermometer employing a thermistor.

Summary. The most accurate method of measuring the humidity of a gas is to determine the dew point. The cooling required for this purpose can be obtained by utilizing the Peltier effect of a combination of certain semiconductors. A description is given of a dew-point hygrometer based on this principle (it is a variant of an instrument using liquid-air cooling, also described). In this apparatus a silver reflector (5 mm thick, surface area 75 mm^2) is cooled to below the dew point. A beam of light is directed from this reflector on to a cadmium-sulphide photoresistor. When the reflector mists over, the beam is diffusely reflected; the photoresistor therefore receives less light and the photocurrent decreases. The latter controls the current through the Peltier element. The control loop thus formed is in equilibrium when the temperature of the reflector is equal to the dew point. The error in the measurement is less than $\frac{1}{2}$ °C.

A TRANSISTORIZED RADIATION MONITOR

by M. van TOL and F. BREGMAN.

621.317.794:621.375.4

This article adds the first measuring instrument to the long list of transistor-equipped devices that have been described in this Review in recent years. The emphasis here lies on the special requirements the circuit of a measuring instrument has to satisfy.

Radiation monitors are simple portable instruments for the detection and measurement of ionizing radiations. They are used in laboratories and hospitals whose staffs work with radioactive substances or with apparatus generating ionizing radiations (X-ray equipment, for example). Another field of application is the tracing of radioactive ores. These simple radiation meters are also suitable for special purposes such as in portable instruments for measuring the level of liquefied gases in cylinders [1]).

Batteries are the obvious form of power supply for portable radiation monitors, and since it is important that their dimensions and weight be small, the designer will be concerned to keep current consumption as small as possible. A radiation monitor which contained only one electronic tube apart from its Geiger-Müller tube, and which in consequence was very economical in operation, was described in this Review in 1953 [2]). It nevertheless required three batteries — a 40 volt anode battery, a 1.5 volt heater battery and a battery for giving the control grid a negative bias of 14 volts. The latter was rendered necessary by the special circuit developed to operate with only one tube.

Small portable electronic devices represent an extremely worthwhile field of application for transistors. In such devices the fullest advantage is taken of the characteristic properties of transistors, namely the small bulk, the low power consumption (and hence low heat dissipation) that results from the lack of heater currents, and the modest voltage requirements that can be met by small batteries, i.e. batteries comprising few cells. A prototype two-transistor radiation monitor, supplied from a single three-volt battery, was in fact developed in the Eindhoven Research Laboratories as early as 1954.

Transistors have less convenient aspects too, these being the spread in characteristics displayed by different individuals of the same type and the sensitiveness of their characteristics to temperature changes. Forming as they do an obstacle to precision, these drawbacks are ordinarily more serious in a measuring instrument than in an audio-amplifier, for example, where no great harm is done if the amplification changes slightly. However, a radiation monitor is a measuring instrument in which the demerits of transistors are not felt so keenly. For one thing, the main requirements are reliability and sensitivity rather than high accuracy; what is more, the measurement in question is based on pulse-counting, and for this purpose one can employ circuits in which the transistor characteristics have little or no influence on the actual reading. It is understandable that reliability should be given first place, for the purpose of the monitor is to measure radiation that is hazardous to human beings. Great sensitivity is called for because in certain cases it may have to detect radiation levels well below the human tolerance level. These requirements can certainly be satisfied just as well with transistors as with tubes.

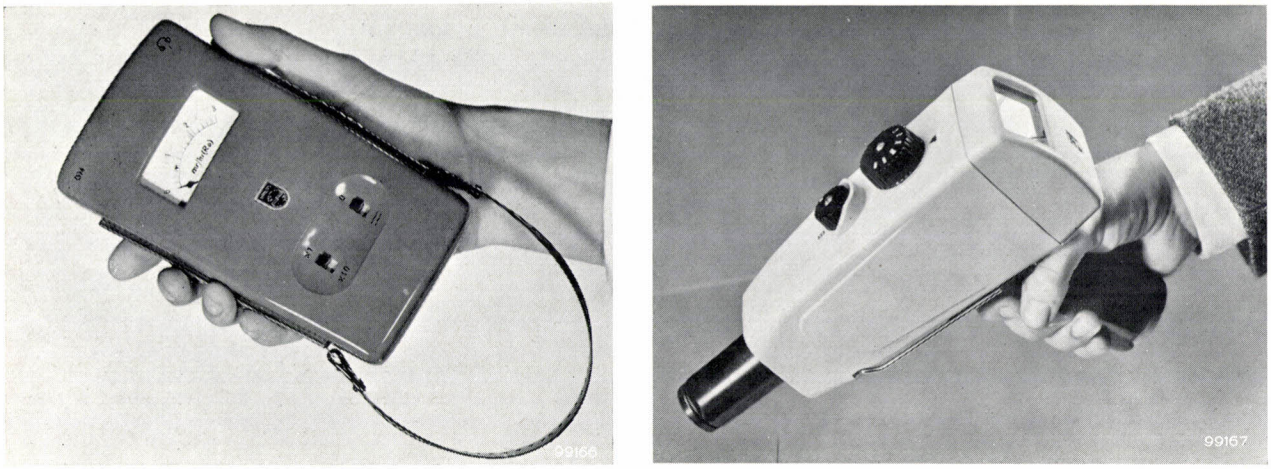
Two commercial versions, types PW 4014 and PW 4012, have finally been evolved from the prototype. Equipped with four and five transistors respectively, the commercial models (illustrated in *fig. 1*) do not differ in any essential respect from the prototype, merely being improvements thereon.

We shall now discuss the circuit of the prototype, given in *fig. 2*. It can be divided into the following parts:

(*a*) H.T. supply circuit.
(*b*) Geiger-Müller tube supplied from this circuit, and delivering electrical pulses when exposed to ionizing radiation.
(*c*) Pulse-shaper that standardizes the height of the pulses.
(*d*) Diode pump-circuit which, supplied with a random sequence of pulses of uniform height, transforms them into a direct current whose value is proportional to the average number of pulses per second. This current is measured with the aid of meter *M*.

The H.T. supply is generated by a simple oscil-

[1]) See for example J. J. Arlman and H. N. L. Hoevenaar, Een niveaudetector voor de praktijk, De Ingenieur 71, Ch. 8 - Ch. 9, 1959 (No. 19) (in Dutch).
[2]) G. Hepp, A battery-operated Geiger-Müller counter, Philips tech. Rev. 14, 369-376, 1952/53.



a *b*

Fig. 1. Two commercial versions of a transistorized monitor for hard γ and β radiation which have been evolved from the prototype developed in the Philips Research Laboratories in 1954 and discussed in the present article.
a) Type PW 4014, pocket radiation monitor.
b) Type PW 4012, pistol model.

lator operating at a frequency of about 10 kc/s, using a transistor T_1, type OC 71. The voltage delivered by the oscillator is stepped up in the transformer and applied to a cascade circuit of three stages, which converts the alternating voltage into a direct voltage of treble its amplitude. The cascade circuit is equipped with selenium diodes.

A direct voltage of the same value could be obtained from a cascade circuit having some other number of stages and a transformer with a different output voltage. Losses in the cascade circuit increase with the number of stages; those in the transformer increase with the transformation ratio. The overall loss curve exhibits a shallow minimum corresponding to the use of three stages, as in the circuit under discussion; however, the particular transformation ratio and number of cascade stages decided upon make hardly any differences to the bulk and weight of the circuit.

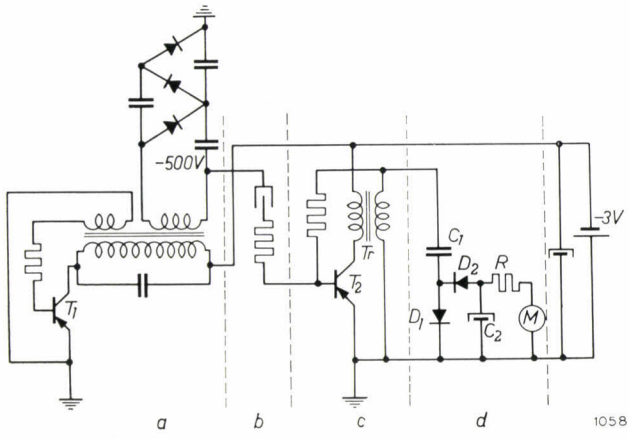


Fig. 2. Basic circuit of the prototype two-transistor radiation monitor. *a*: high-tension supply circuit for Geiger-Müller tube. *b*: G.M. tube with series resistor. *c*: pulse-shaper. *d*: diode pump-circuit with milliammeter *M*. The circuits of the commercial versions appearing in fig. 1 do not differ in essentials from that shown here.

The G.M. tube is of type 18 503 or 18 504; in both types the quenching gas is a halogen. On account of their robustness and almost unlimited life, these types of tube are particularly suitable for use in an instrument whose prime requirement is dependability. Moreover, their characteristics exhibit a long "plateau" with a particularly gentle slope (*fig. 3*). The high tension across the tube can therefore vary within wide limits (370 to 650 V) without greatly affecting the rate at which the tube delivers pulses when exposed to radiation of constant intensity [3]). For this reason it is possible to use a high-tension supply circuit of such simple design; for the same reason, there is no need to pay overmuch attention to the characteristics of transistor T_1. Provided the oscillator oscillates — and that is not asking a great deal — a usable high-tension supply will be available.

The voltage supplied is about 500 V, and the average current taken by the G.M. tube does not exceed 30 μA. The maximum power required is therefore 15 mW. A three-volt battery supplies the oscillator (and the other circuits). The high-tension circuit has an efficiency of about 60% and therefore takes less than 0.01 A from the battery — a very low current for a rod battery.

[3]) K. van Duuren, A. J. M. Jaspers and J. Hermsen, G. M. counters, Nucleonics **17**, No. 6, 86-94, 1959.

As already stated, the value of the high-tension voltage has little effect on the *rate* at which the G.M. tube delivers pulses. It does, however, have an effect on the *height* of the pulses. Since there is no stabilization of the H.T. supply, the average current through the G.M. tube cannot be used directly as a measure of the count rate. For this reason the pulses are fed to the pulse-shaping circuit *c* (see fig. 2), which consists of a type OC 71 transistor T_2 and a transformer *Tr*. Circuit *c* is a blocking oscillator [4]). T_2 acts as a switch that closes each time the G.M. tube delivers a pulse. When this happens, the primary winding of *Tr* carries the full battery voltage (3 V). The secondary winding of *Tr* thus has induced in it a voltage pulse whose amplitude is related to the battery voltage by the ratio of the numbers of turns. The time that T_2 remains in the conductive state, and hence also the width of the voltage pulse in the secondary winding of *Tr*, depend

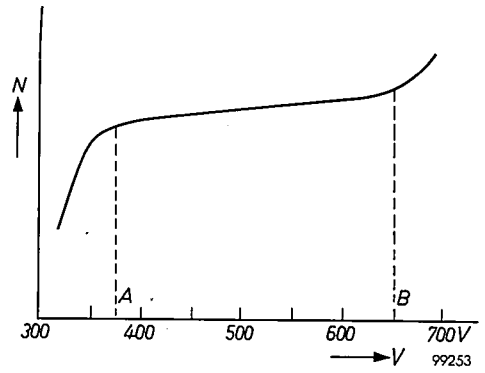


Fig. 3. Count-rate characteristic of a type 18 504 (halogen-quenched) Geiger-Müller tube. *N*, the number of pulses the tube delivers per second when exposed to a constant radiation intensity, is plotted as a function of the voltage *V* across the tube. The characteristic has a particularly low plateau slope (the "plateau" is the part of the curve between *A* and *B*).

amongst other things on the characteristics of the transistor. These also have an influence on the slope of the pulse edges, but the height of the pulses depends on the battery voltage alone.

A diode pump-circuit *d* measures the rate at which the voltage pulses, now of standard height, are produced. The circuit owes its name to the analogy with a piston force pump, in particular as used for compressing a gas. Diodes D_1 and D_2 represent the suction and delivery valves respectively; storage capacitor C_1 may be compared with the barrel of the pump, and the much larger buffer capacitor C_2 with the tank into which gas (i.e. electrons) is being pumped. However, charge continually drains away from the latter via resistor *R* and ammeter *M*.

A characteristic property of the circuit is that the current through *M* is dependent on the number and the height of the pulses entering it but not — provided they are sufficiently wide, allowing C_1 to be fully charged — on their duration or the slope of their edges. (Compare with the piston force pump, which displaces a quantity of gas that is dependent on the number of strokes per sec and on the pressure of the gas entering the pump, but not on the manner in which the barrel fills up, provided only that it fills with gas up to the full inlet pressure.) The diode pump-circuit is therefore very suitable for handling the pulses of uniform height that are delivered by the pulse-shaper *c*. It is possible, by a correct choice of values for C_1, C_2 and *R*, to make the current through *M* proportional, or nearly so, to the pulse rate. A recent article in this Review [5]) may be referred to for a more detailed discussion of the way in which a diode pump-circuit functions.

We have seen above that the constituent parts of the monitor have been adapted to one another with the object of getting a meter reading that is independent of transistor characteristics. We may recapitulate as follows.

The H.T. voltage yielded by circuit *a* (fig. 2) depends on T_1. It is applied to G.M. tube *b*, which delivers pulses at a rate which is independent of the voltage applied and hence of T_1. The height of the pulses *is* dependent on T_1. Pulse-shaper *c* gives these pulses a constant *height*. Their *shape* (i.e. their duration and the slope of their edges) depends, however, on the characteristics of T_2; the influence of T_1 has been eliminated. Diode pump-circuit *d* converts these pulses of standard *height* into a current that is independent of their *shape*. Thus the influence of T_2 has also been eliminated.

Even so, the current through the measuring instrument is still dependent on the battery voltage, for the height of the pulses delivered by circuit *c* is proportional to that voltage (see above). Facilities for checking the battery voltage have therefore been provided in the commercial versions in fig. 1, the indication being given on the meter (*M*).

The ranges measured by pocket radiation monitor PW 4014 are 0 to 3 mr/h (milliröntgen per hour) and 0 to 30 mr/h [6]). The pistol model, PW 4012, has the ranges 0 to 1 mr/h, 0 to 10 mr/h and 0 to 100 mr/h. The design of both models allows for the

[4]) For the functioning of a blocking oscillator see, for example, B. Chance *et al.*, Waveforms, No. 19 of the Radiation Laboratory Series, McGraw-Hill, New York 1949.

[5]) J. J. van Zolingen, Philips tech. Rev. **21**, 134-144, 1959/60 (No. 4/5), in particular p. 138 et seq.

[6]) Röntgen, röntgen per hour and other units relating to ionizing radiation were discussed in N. Warmoltz and P. P. M. Schampers, A pocket dosemeter with built-in charger, for X-radiation and gamma radiation, Philips tech. Rev. **16**, 134-139, 1954/55.

connection of a type 18 509 counter tube, which extends the PW 4014 ranges to 0-60 and 0-600 mr/h, and the PW 4012 ranges to 0-20 and 0-200 mr/h. To give some idea of the significance of these figures, we would add that 5 röntgens per year is regarded as an acceptable dose for persons exposed to radiation by reason of their occupation. In a working year of 2000 hours this amounts to 2.5 mr/h. The tolerance dose for other classes of the population is lower.

A circuit giving a new kind of meter scale

A new circuit has meanwhile been developed in the laboratory in which the scale of the meter is progressively compressed (though *not* logarithmically) at higher pulse rates. This means that, as for a logarithmic scale, more economical use is made of the available scale length than when a linear scale is used. The deflections obtained with this circuit differ from those of a logarithmic scale in that zero deflection corresponds to absence of radiation and full-scale deflection to an infinitely high radiation level. Consequently the needle never goes beyond the extremes of the scale — a very convenient feature in a radiation monitor: the user always has immediate indication of the radiation level without range-switching.

The principle of the new circuit is as follows. *Fig. 4* shows a flip-flop circuit with two transistors T_3 and T_4. Negative pulses from the G.M. tube arrive at terminal A at an average rate of N per second. After such a negative pulse on A the circuit is left in the stable state in which T_3 conducts; suppose

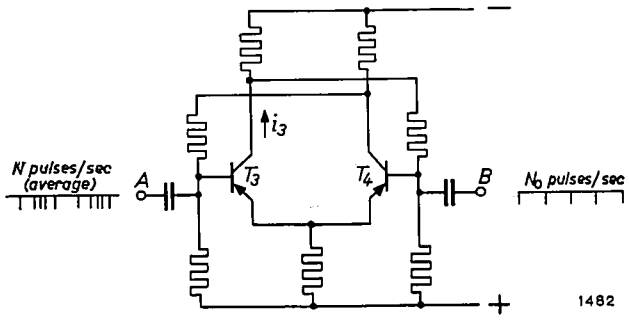


Fig. 4. Flip-flop circuit in which terminal A goes to the G.M. tube, and terminal B to a pulse generator that delivers pulses at a constant rate of N_0 per sec. The mean current through transistor T_3 is a measure for N, the average number of pulses delivered by the G.M. tube per sec.

that the current i_3 then flowing through T_3 is I_3. A pulse generator supplying negative pulses of constant repetition frequency N_0 is connected to terminal B. After each of these pulses the circuit is left in the state in which T_3 is non-conducting. The mean current through T_3 is now a measure for N; this can be seen with the help of *figs. 5a* and *b*. In each diagram the pulses arriving at A, those

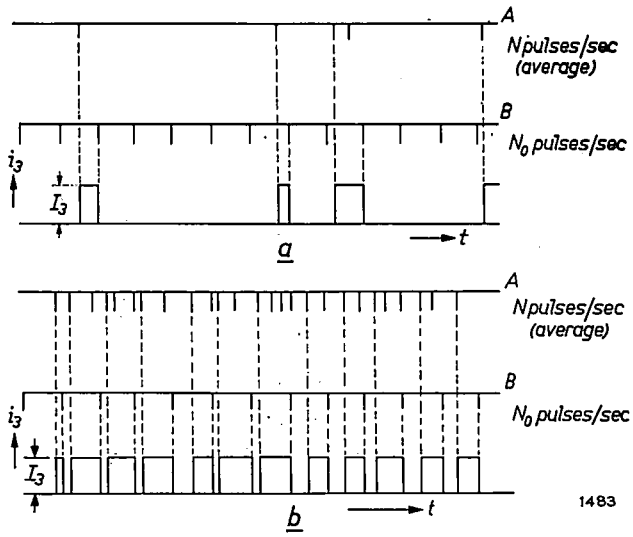


Fig. 5. Diagram to explain the functioning of the circuit in fig. 4.

arriving at B and the current through T_3 are plotted one below the other as functions of time; the current i_3 through T_3 can only have one of two values, zero or I_3. When T_3 has been rendered conducting by a pulse from the G.M. tube, a current I_3 flows until stopped by the next pulse of the regular train from the generator. It will be seen from fig. 5*a* that if N is much less than N_0, as in that diagram, then the mean current through T_3 will have a low value. Indeed, if the G.M. tube stops delivering pulses altogether, the mean current will be zero. Fig. 5*b* shows the situation when N greatly exceeds N_0. Here T_3 is conducting almost all the time. At very high radiation levels the mean current through the transistor has a value approaching I_3.

The two extreme cases $N \ll N_0$ and $N \gg N_0$ lend themselves to a simple quantitative treatment, as follows.

For the case $N \ll N_0$, almost every one of the N pulses delivered by the G.M. tube per sec will reach T_3 when it is in the cut-off state. At any instant the average time that will elapse before the next pulse arrives from the generator is $\frac{1}{2}(1/N_0)$ sec. Roughly, then, T_3 conducts for a proportion of the time given by $\frac{1}{2}N/N_0$, and the mean current through it is $\frac{1}{2}(N/N_0)I_3$. The lower end of the scale therefore has a linear character.

The case $N \gg N_0$ can be dealt with analogously. Of the N_0 pulses delivered by the generator each second, almost every one reaches T_3 when it is in

the conducting state. At any instant the average time required for the next pulse to arrive from the G.M. tube is $1/N$ sec. (The factor $\frac{1}{2}$, appropriate to pulses arriving in a *regular* train, is absent here on account of the statistical distribution of the pulses delivered by the G.M. tube.) Roughly, then, T_3 is cut off for a proportion of the time given by N_0/N, and the mean current through it is $(1-N_0/N)I_3$. At the upper end of the scale, therefore, the deflection of the needle approaches the full-scale value (I_3) hyperbolically.

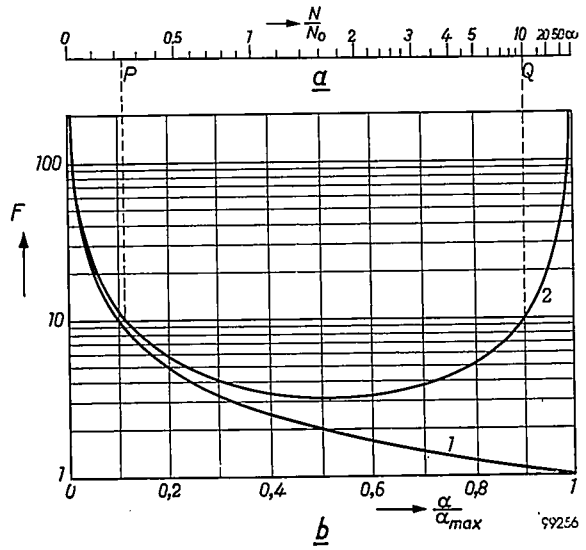


Fig. 6. *a*) Scale obtained with the new circuit; it provides readings of N/N_0, the ratio between the average number of pulses delivered by the G.M. tube per sec and the constant number of pulses delivered by the pulse generator per sec. *b*) The quantity F indicates the percentage measuring error arising from a (constant) deflection error of 1% of α_{max}, the full-scale angular deflection. F is plotted logarithmically as a function of α/α_{max} for a linear scale (curve *1*) and for the new scale (curve *2*).

The general formula for the mean current $\bar{i}_3$ through T_3 is:

$$\bar{i}_3 = I_3\left(1 - \frac{1 - e^{-N/N_0}}{N/N_0}\right).$$

A scale in terms of the ratio N/N_0, based on the above formula, is drawn in *fig. 6a*. If N_0 is made equal to the G.M. pulse rate corresponding to 1 mr/h (this rate depends on the properties of the tube employed), then we obtain a scale whose unit division corresponds to 1 mr/h.

Comparison of the new scale with a linear scale makes clear how economical the former is in its use of the available scale length. The most obvious basis for comparing two meter scales, both of which are used for measuring a quantity x, is the relative error of measurement, $(\Delta x)/x$. This error in x proceeds from the error $\Delta\alpha$ in the deflection α undergone by the needle, and is given by

$$\frac{\Delta x}{x} = \frac{1}{x}\frac{dx}{d\alpha}\Delta\alpha.$$

Hence the relative error in x per unit deflection error is equal to $(1/x)(dx/d\alpha)$; this expression, plotted as a function of α, is therefore a suitable yardstick for comparing the merits of different scales. In fig. 6*b* the dimensionless quantity

$$F = \frac{\alpha_{max}}{x}\frac{dx}{d\alpha}$$

is plotted as a function of another dimensionless quantity α/α_{max} to give curves that are independent of the unit in which α is expressed. α_{max} stands for the full-scale deflection. F denotes the percentage measuring error proceeding from a deflection error $\Delta\alpha$ that is 1% of α_{max}. If it is stipulated that F shall nowhere exceed 10, then it will be clear from fig. 6*b* that the linear scale has a usable range extending from $\alpha/\alpha_{max} = 0.1$ to $\alpha/\alpha_{max} = 1$, while that of the new scale extends from P to Q, i.e. from $\alpha/\alpha_{max} \approx 0.11$ to $\alpha/\alpha_{max} \approx 0.9$. The usable portion of a linear scale covers a measuring range whose upper limit occurs at a value 10 times greater than the value read at its lower limit. On the new scale in fig. 6*a* points P and Q occur at values of approximately 0.233 and 10 respectively; the upper limit thus occurs at a value 10/0.233 times or a good 40 times greater than the value read at the lower limit. *Fig. 7a* shows two linear scales covering adjoining ranges; together they cover the range from 0.3 to 30 mr/h with a percentage measuring error F nowhere exceeding 10. Two scales of the new type covering adjoining ranges, so chosen that the lower useful limit of the top scale again lies at 0.3 mr/h, are reproduced in fig. 7*b*; with the same maximum F value of 10, these have an overall range going up to 500 mr/h. A reasonable estimate can moreover be made up to 2000 mr/h.

Calibration of the new scale is particularly easy. The flip-flop circuit is held first in one and then

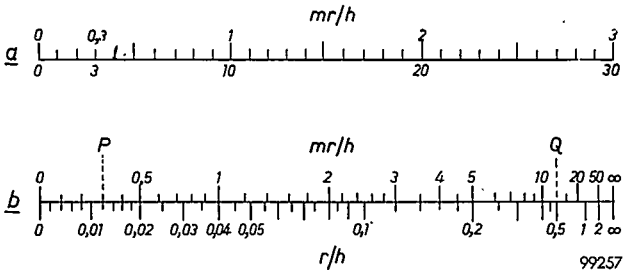


Fig. 7. *a*) Two linear scales which together cover the range from 0.3 to 30 mr/h with a percentage measuring error $F \leqslant 10$. *b*) Two complementary scales of the new type. Here the total range covered, over which $F \leqslant 10$, extends from 0.3 mr/h (as on the linear scale in *a*) to 500 mr/h (0.5 r/h).

in the other of its stable states; the circuit is adjusted so that the meter reads zero in the state in which T_3 is cut off and adjusted to give full-scale deflection in the state in which T_3 conducts. All that is now necessary is to give N_0 the desired value. The normal practice will be to place the monitor in the radiation field of a standard radioactive preparation and then to adjust the generator to give a pulse repetition frequency such that the correct meter reading is obtained. Thus only one point has to be calibrated on each measuring range, as in the case of a linear scale. The new scale is much more convenient to calibrate than a logarithmic scale. Owing to the lack of a zero, a logarithmic scale has to be calibrated at two points, and adjustment at one point entails readjustment at the other.

Summary. Portable radiation monitors are an obvious case in which transistors can be used to advantage. As measuring instruments, however, the circuit employed must be so designed that the readings are insensitive to the transistor characteristics (which may shift considerably with changes of temperature). A circuit of this kind is described. The authors also discuss a method of arriving at a (non-linear) scale that covers the whole range between zero radiation (zero deflection) and an infinitely high radiation level (full-scale deflection). Readings are sufficiently accurate (less than 10% error for a deflection error equivalent to 1% of the scale length) over a part of the scale extending from 0.1 to 0.9 of the scale length. The radiation-level reading at the upper limit of this range is a good 40 times higher than that at the lower limit. The instrument can easily be adjusted to bring any given radiation level within the usable range. The calibration procedure is particularly straightforward.

STRAY CAPACITANCES IN NEON INSTALLATIONS

by J. J. WILTING.

621.327.42

High-voltage neon-filled discharge tubes, used widely as luminous signs, can generate strong electrical transients. These are bound up with the stray capacitances in the installation. The article below deals with the favourable as well as the unfavourable consequences of these phenomena.

A gas discharge which is maintained by an alternating current of low frequency is periodically extinguished and must therefore be periodically reignited (twice in every cycle). The voltage needed for restarting the discharge depends on the nature and the pressure of the gas filling and on the geometry of the tube. In the case of long discharge tubes, as frequently used for luminous signs, the reignition voltage U_i is not much higher than the burning voltage U_b if the gas filling is mercury vapour (*fig. 1a*). With neon gas, on the other hand, U_i may well be three or four times higher than U_b (fig. 1*b*).

The high reignition voltage of neon tubes thus calls for a supply transformer with a particularly high no-load voltage, and is therefore quite expensive; on the other hand, an upper limit is set to the permissible transformer voltage by safety regulations. The presence of the stray capacitances of the transformer and of the high-tension cable connecting the tubes to the transformer is fortunate in this respect, in that it offers the possibility of generating a high voltage pulse at approximately the moment when reignition is needed. This facilitates the reignition and lowers the necessary no-load voltage of the transformer.

As the current periodically drops to zero (initiating the "dark period", i.e. a currentless interval lasting until the reignition), a transient is produced in the electrical circuit. This consists of a damped oscillating voltage superimposed on the voltage across the neon tubes, with a frequency of about 500 to 1200 c/s, which is thus substantially higher than the mains frequency. The sum of this damped oscillation (the above-mentioned voltage pulse) and the voltage of the temporarily unloaded transformer may be high enough for reigniting the neon tubes.

Capacitance (and stray capacitance is always present) is essential to the occurrence of this useful effect. The other side of the picture, however, is that the capacitance, after charging up to a high voltage, immediately discharges through the neon tubes after every reignition. The discharge takes the form of a strong current surge which makes the neon tubes burn unsteadily and shortens their life.

In the following we shall consider the means of promoting the generation of an effective voltage pulse whilst limiting the current surges to tolerable values.

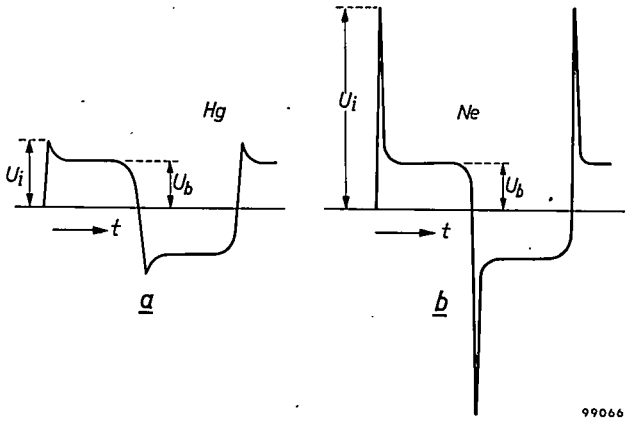


Fig. 1. *a*) In long discharge tubes containing mercury vapour (under low pressure) the reignition voltage U_i is not much higher than the burning voltage U_b.
b) In long neon tubes U_i is several times higher than U_b.

The voltage pulse for reignition

Transformers for neon tubes (one or more tubes in series) are generally designed with some magnetic-flux leakage, with the object of obtaining a specific internal reactance that will take up the difference between the no-load and burning voltages. Some designs in common use are shown in *fig. 2*. By means of magnetic shunts *Sh* the reactance is adjusted such that the current assumes its rated value. The advantage of the balanced designs (fig. 2*b* and *d*) over the unbalanced is that the secondary voltage may permissibly be twice as high as the maximum voltage with respect to earth as prescribed by safety regulations.

Since the tubes become conducting and non-conducting fairly abruptly, they may be considered, as regards their circuit behaviour, as a switch in series with (principally) a resistance: the switch opens at the moment $t = 0$ when the current drops to zero,

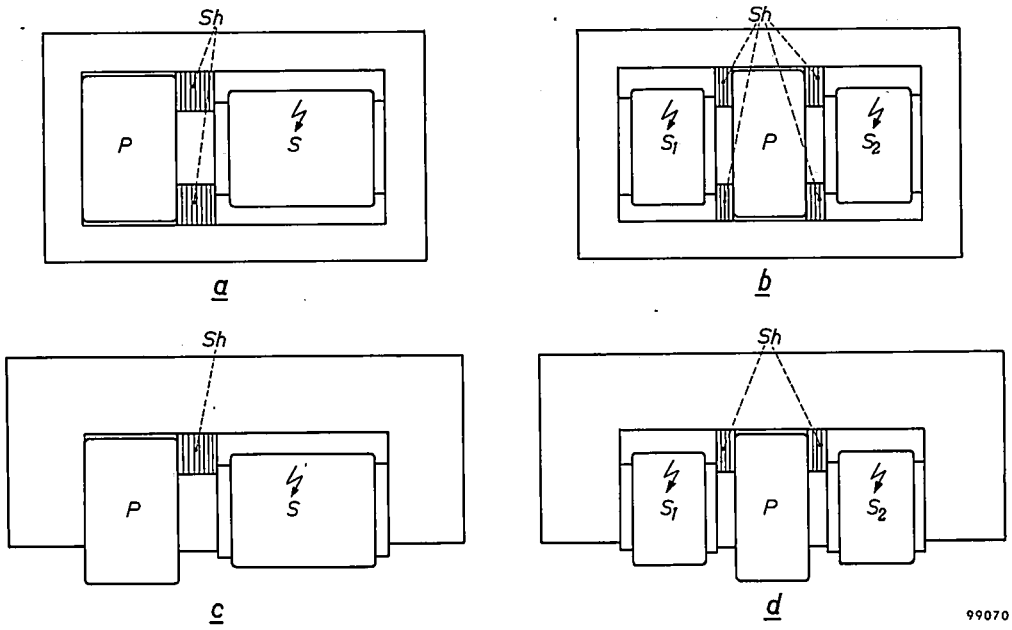


Fig. 2. Common designs of high-tension transformers for neon tubes. P primary coil. S secondary coil (in b and d divided into balanced halves, S_1 and S_2). a and b have cores of the shell type; c and d of the yoke type. Sh magnetic shunts for adjusting the leakage reactance to a specific value.

and closes again at the moment of reignition. We denote the no-load voltage by $e = E \sin(\omega t + \varphi)$; the voltage across the neon tubes at $t = 0$ (beginning of the dark period) is the superposition of the instantaneous value $e_0 = E \sin \varphi$ of the no-load voltage at $t = 0$ and the transient voltage Δe (*fig. 3*).

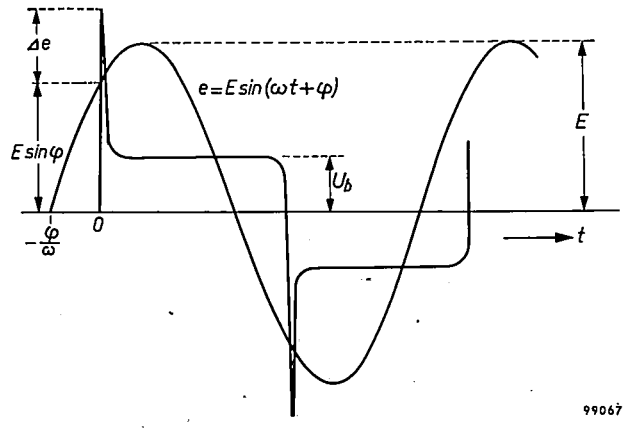


Fig. 3. Voltage waveform in a burning neon tube. No-load voltage of transformer: $e = E \sin(\omega t + \varphi)$. At the moment at which the current drops to zero (here also the moment of reignition) $t = 0$. The no-load voltage at that moment is $E \sin \varphi$; but superposed on this is a voltage Δe, supplied by the transient, which must be sufficient to ensure reignition. U_b = burning voltage.

The transient can be calculated to a good approximation on the basis of a simplified equivalent circuit. *Fig. 4a* shows the diagram of a balanced neon installation with the stray capacitances C_t of the transformer and C_k of the high-tension cable. Trans-

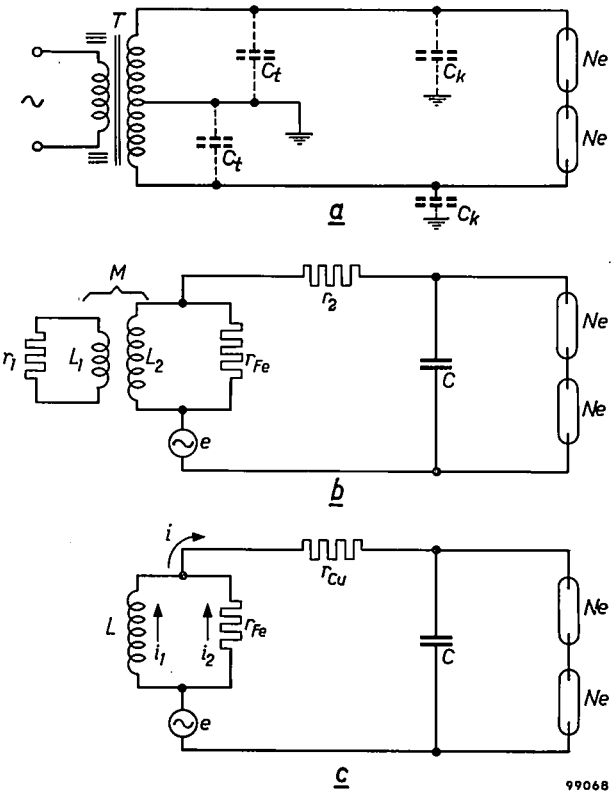


Fig. 4. *a*) Balanced transformer of the type in fig. 2*b* or *d*, supplying two neon tubes in series. C_t and C_k are stray capacitances of transformer and high-tension cable, respectively. *b*) Equivalent circuit of (*a*) with AC voltage source e placed in the secondary circuit. L_1 and r_1, L_2 and r_2 represent the inductance and resistance of the primary and secondary coils, respectively. M mutual inductance. r_{Fe} effective resistance due to iron losses. C total stray capacitance.
c) Simplified equivalent circuit. L leakage inductance. r_{Cu} effective resistance due to all copper losses.

ferring the voltage source to the secondary circuit and short-circuiting the primary coil produces the equivalent circuit in fig. 4*b*. All stray capacitances are now lumped together in the capacitance C. The resistance and inductance of the primary coil are, respectively, r_1 and L_1, and of the secondary coil r_2 and L_2. A resistance r_{Fe} is added to allow for the iron losses.

By further simplification, this circuit can be reduced to that of fig. 4*c*. Here the inductance L is given by

$$L = (1 - k^2)\, L_2,$$

where k is the coupling coefficient of the transformer. To a first approximation L and r_{Fe} may be regarded as constants.

The resistance r_{Cu} in fig. 4*c* allows for the total copper losses:

$$r_{Cu} = r_2 + \frac{L_2}{k^2 L_1}\, r_1.$$

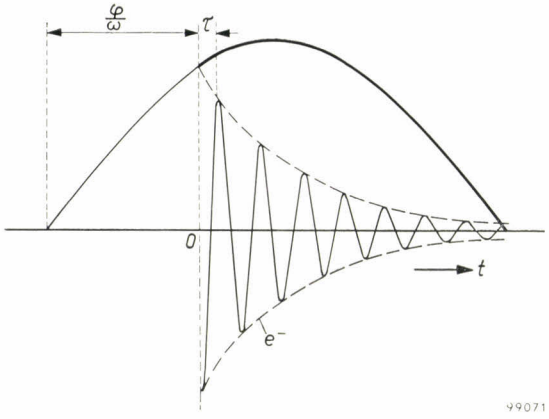


Fig. 5. The sine wave represents the no-load voltage, the damped oscillation is the transient.

Making certain minor approximations, we can derive for fig. 4*c* the following equation describing the variation of the capacitor voltage u with time t:

$$u(t) = E \sin(\omega t + \varphi) - E \sin\varphi \,.\, e^{-\alpha t} \cos \beta t. \quad (1)$$

In this expression ω is the angular frequency of the mains, α the damping constant and β the natural angular frequency of the excited "high-frequency" resonant circuit L-C; see *fig. 5*. For α and β we can write:

$$\alpha \approx \frac{r_{Cu}}{2L} + \frac{1}{2 r_{Fe} C} \quad \ldots\ldots \quad (2)$$

and

$$\beta \approx \frac{1}{\sqrt{LC}}. \quad \ldots\ldots\ldots \quad (3)$$

An oscillogram of the transient is shown in *fig. 6*.

Equations (1), (2) and (3) are obtained as follows. Let the currents through r_{Cu}, L and r_{Fe} in fig. 4*c* be i, i_1 and i_2, respectively; then:

$$i = i_1 + i_2,$$

$$L \frac{di_2}{dt} = i_2 r_{Fe},$$

$$i_2 r_{Fe} + i r_{Cu} + u = e,$$

$$i = C \frac{du}{dt}.$$

Eliminating i, i_1 and i_2 from these expressions, we find a differential equation of the form

$$\frac{d^2u}{dt^2} + a \frac{du}{dt} + bu = ce + d \frac{de}{dt}, \quad \ldots \quad (4)$$

where a, b, c and d are certain functions of L, C, r_{Cu} and r_{Fe}. For the overshoot above the no-load voltage the periodic solutions of the differential equation (4) are the only ones of importance. These are of the form

$$u(t) = A \sin(\omega t + \gamma) + B\, e^{-\alpha t} \sin(\beta t + \delta). \quad . \quad (5)$$

This equation expresses the superposition of the no-load voltage of angular frequency ω and a damped oscillation of higher angular frequency β. The coefficient A and the phase angle γ follow from an easily found particular solution. If the initial conditions are known, B and δ can also be calculated.

The initial moment is the moment at which the current i reaches the zero value. The burning voltage then also falls to zero, so that we may write:

$$u(0) = 0.$$

From (5) it therefore follows that

$$A \sin\gamma + B \sin\delta = 0. \quad \ldots\ldots \quad (6)$$

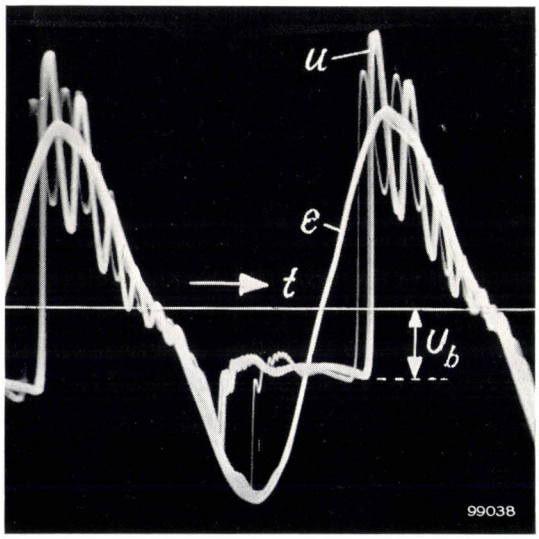


Fig. 6. Oscillogram of transformer voltage e (no-load operation) and of the voltage u across a neon tube which passes current in one direction only, a rectifier being connected in series with it to make the oscillogram clearer. The transformer voltage alone, without the transient, is too low to guarantee reignition of the neon tube in every cycle. The superposition of the two curves in fig. 5 can be recognized in the waveform of u.

The second initial condition is given by the magnitude of du/dt at the moment $t = 0$. This depends on various conditions. Suppose

$$\left(\frac{du}{dt}\right)_{t=0} = U_0',$$

then from (5) we find:

$$U_0' = A\,\omega\cos\gamma + B\,\beta\cos\delta - B\,\alpha\sin\delta. \quad . \quad . \quad (7)$$

From (6) and (7) we now obtain:

$$\tan\delta = -\frac{A\beta\sin\gamma}{A\,\omega\cos\gamma + A\,\alpha\sin\gamma - U_0'}.$$

Since the effect of the resistances r_{Cu} and r_{Fe} is negligible compared with that of the inductance L and the capacitance C, it follows from fig. 4c that, approximately,

$$A \approx \frac{E}{1-\omega^2 LC} \quad \text{and} \quad \gamma \approx \varphi.$$

Where the high-tension cables are not unduly long, $1/\omega C \gg \omega L$, hence $A \approx E$. With this approximation equation (5) becomes

$$u(t) \approx E\sin(\omega t + \varphi) - E\frac{\sin\varphi}{\sin\delta}\,e^{-\alpha t}\sin(\beta t + \delta). \quad . \quad (8)$$

Measurements have shown that, at all encountered values of U_0', the value of δ is only a few degrees smaller than 90°, so that $\sin\delta \approx 1$. Insight into the phenomena taking place can therefore be gained by putting the angle δ in (8) equal to $\frac{1}{2}\pi$

$$u(t) \approx E\sin(\omega t + \varphi) - E\sin\varphi\,.\,e^{-\alpha t}\cos\beta t,$$

and this is the above equation (1).

Finally, the damping factor α and the angular frequency β of the damped oscillation follow from (5) after writing A and B in terms of the constants a, b, c and d, and substituting for the latter in terms of the various circuit parameters:

$$\alpha = \frac{L + r_{Cu}\,r_{Fe}\,C}{2(r_{Cu} + r_{Fe})LC}$$

and

$$\beta = \sqrt{\frac{r_{Fe}}{(r_{Cu} + r_{Fe})LC} - \alpha^2}.$$

By approximation these expressions can be simplified to formulae (2) and (3) given above.

The required reignition voltage

The reignition voltage of a given neon tube depends principally on the time interval τ between extinction and reignition, and, where the tube is reignited by a voltage *pulse*, on the duration of this pulse. (The magnitude of the current through the tube before the extinction also plays some part, but this need not be considered here.)

The first of the two effects — the dependence on τ — is demonstrated in *fig. 7a* by curves *1* and *2*. The curves represent the reignition voltage as a function of τ measured on two neon tubes of the same diameter (13 mm), but measuring respectively 3.85 m and 1.85 m in length. The measurements were done with direct voltage which was switched off (electronically) for a variable time τ. Where τ is of the order of some microseconds the reignition voltage is approximately equal to the burning voltage. When τ is increased, the reignition voltage also increases until it reaches a maximum at $\tau \approx 1000$ μsec, where it is equal to the (initial) ignition voltage.

If the tube operates on alternating voltage, it is reignited by a short-lived peak and the situation is considerably less favourable. Because of the short duration of the pulse, the discharge can only be properly initiated if the peak voltage is much higher than the direct voltage that was sufficient for the measurements mentioned above. This situation is represented in fig. 7*b*. The lines *1* and *2* roughly correspond to curves *1* and *2* in fig. 7*a*, and thus

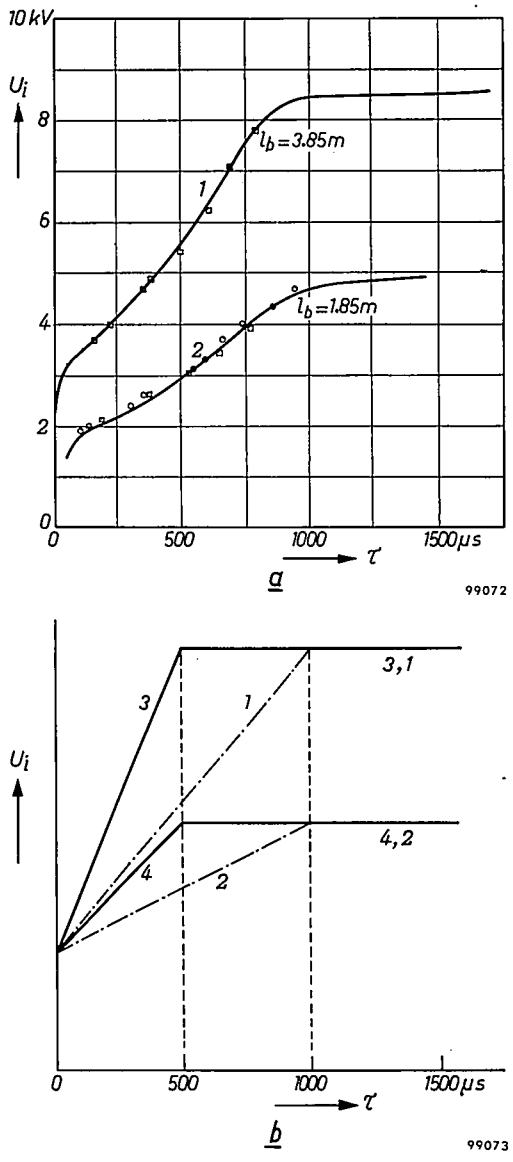


Fig. 7. *a*) Reignition voltage U_i for neon tubes (13 mm inside diameter) operated on DC, the supply being electronically switched off for a variable interval τ, as a function of τ. Curve *1* refers to a tube 3.85 m long, curve *2* to a tube 1.85 m long, both on a current of 25 mA.
b) Reignition voltage on AC operation (schematic); *3* refers to the 3.85 m tube, *4* to the 1.85 m tube. For comparison the DC lines *1* and *2* are shown, which roughly correspond to *1* and *2* in (*a*).

apply to direct voltage. In the case of alternating voltage having a waveform as shown in fig. 6 and a pulse duration of $\frac{1}{2}\tau$, the curves are as shown by *3* and *4* for neon tubes of lengths 3.85 and 1.85 m respectively: the reignition voltage has now risen to the ignition voltage at $\tau \approx 500$ μsec instead of at 1000 μsec.

In a normally burning neon tube the dark period is longest when the first positive peak of the damped oscillation only just reaches the reignition voltage. At the moment of reignition, a complete half-cycle of the damped oscillation has then elapsed since the moment $t = 0$ (fig. 5). As the oscillation has the angular frequency β, given by equation (3), the maximum dark period τ_{max} is:

$$\tau_{max} = \frac{\pi}{\beta} = \pi\sqrt{LC}.$$

In large neon installations (no-load voltage 8 kV) the values of L and C are generally such that $\tau_{max} > 500$ μsec. Neon tubes of 13 mm diameter (for which fig. 7 holds) therefore require the full ignition voltage for reignition in such an installation. At larger tube diameters the curves *3* and *4* in fig. 7*b* shift to the right; it may then well be, particularly in smaller installations with higher β, that the working point for reignition comes to lie on the sloping portion of the curve.

Relation of maximum tube length to transformer voltage

A factor of considerable practical importance in neon installations is the maximum length of tube that will operate efficiently on a given transformer voltage. This length is sharply limited. The longer the tube, the higher is the *required* reignition voltage, but the lower is the *available* reignition voltage. The truth of the latter statement will now be demonstrated below.

The first voltage peak $u(t)$ occurs at about the moment $t = \tau = \pi/\beta$. Substituting this value for t in equation (1), we obtain:

$$u(\tau) \approx E \sin\left(\frac{\omega}{\beta}\pi + \varphi\right) + E \sin\varphi \exp\left(-\frac{\alpha}{\beta}\pi\right). \quad (9)$$

If the tube length is changed, the current is restored to the rated value by adjusting the magnetic shunts. This is attended by a change in the phase angle φ. The reason is that an increase in tube length represents increased resistance in the circuit, and thus necessitates adjustment of the shunts to obtain a smaller leakage reactance. Both changes (larger resistance, smaller inductance) cause a reduction of φ, and hence a reduction of both terms on the right-hand side of equation (9) (the argument of the first term, $(\omega\pi/\beta) + \varphi$, is an angle in the *first* quadrant). These terms together constitute the available reignition voltage.

On the basis of (9) we shall now examine the measures that can be taken to keep the available reignition voltage adequate for as long as possible when increasing the length of the tube. The parameters that can be manipulated in this equation are the angular frequency β and the damping factor α. It is seen from (9) that the first term of the right-hand side increases when β is decreased, and that the second term increases when α/β is decreased. We thus require a reduction of β accompanied by a relatively greater reduction of α.

Let us first consider the possibilities of reducing the angular frequency β. This can be done by increasing C or L, or both. C is governed mainly by the length of the high-tension cable. The magnitude of C, hence effectively the length of the cable, is limited, however, because the surge currents increase with C; if they are excessive, they cause the tubes to burn unsteadily and shorten their life (see next section). The practical limit is reached with cables of a few metres long. As regards the inductance L, it should be noted that this quantity is not identical with the leakage inductance that determines the steady-state current. The rapid flux changes that give rise to the transient are superposed on the steady-state alternating field. The inductance L concerned with the transient is therefore determined by the incremental permeability μ_i of the type of laminated core steel employed; μ_i depends on the pre-magnetization due to the main magnetic field during the dark period. In designing the transformer one can therefore try to make L large (and hence β small) by using lamination steel having a high μ_i and by not saturating the core too strongly.

Let us now consider the ratio α/β, which should also be small. We have seen that the damping factor α is given by:

$$\alpha \approx \frac{r_{Cu}}{2L} + \frac{1}{2r_{Fe}C}. \quad \ldots\ldots \quad (2)$$

Inserting the numerical values encountered in practice we find that the second term, $1/(2r_{Fe}C)$, is predominant. At a given C, therefore, α is almost solely dependent on the iron losses. *Fig. 8* shows three curves of α plotted as a function of β for a transformer of the type 2×4000 V, 25 mA. As might be expected, α decreases when the iron losses are reduced by using either thinner laminations or a steel with a higher silicon content. It is also seen from fig. 8 that α/β decreases with decreasing

frequency; a lower β does in fact therefore increase the second as well as the first term on the right-hand side of (9).

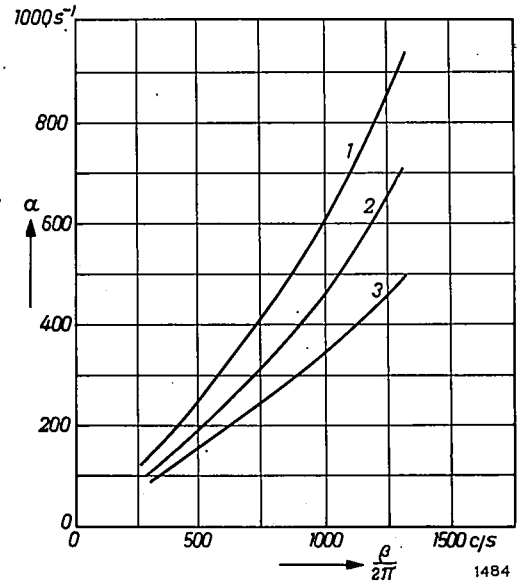


Fig. 8. Damping factor α as a function of frequency $\beta/2\pi$ of the transient, for various types of silicon steel used for the core laminations of a 2 × 4 kV, 25 mA transformer. Curve *1*: low Si content, laminations 0.5 mm thick. Curves *2* and *3*: higher Si content, laminations respectively 0.5 mm and 0.35 mm thick.

For the above type of transformer, loaded with several neon tubes in series (13 mm diameter, total length 6 m), *fig. 9* shows the measured "critical mains voltage" in relation to the nominal mains voltage as a function of the total stray capacitance C. (By critical mains voltage is meant the lowest voltage at which the tubes burn steadily, that is without flickering; it is thus the mains voltage at which the generated reignition voltage is only just sufficient.) The dashed curve was calculated from

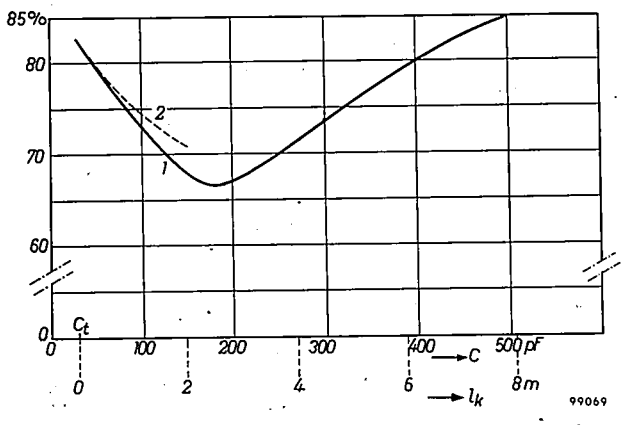


Fig. 9. Ratio of critical to nominal mains voltage, as a function of the stray capacitance C. Curve *1*: measured. Curve *2*: calculated. Of the stray capacitance 30 pF was contributed by the transformer; the rest was due to the cable, whose lengths l_k correspond to the abscissa values. Transformer 2×4 kV, 25 mA.

equation (9). Its deviation from the measured curve is mainly attributable to the non-linearity of the inductance L. The fact that the measured curve rises again on the right is due to the strong current surges affecting the steady burning of the tube; we shall return to this subject presently.

Since the mains voltage may be lower than the nominal value, and since the ignition voltage may increase during the life of neon tubes, the requirement for *new* tubes in the most *unfavourable* arrangement (with very short connections) is that the critical mains voltage should not exceed 80% of the nominal value.

Current surges

At the moment of reignition the stray capacitance C, which is in parallel with the neon tubes, has the voltage $u(\tau)$ across it; see equation (9). The reignition makes the tubes abruptly conducting, causing the sudden discharge of the capacitance. The resultant current pulse (*fig. 10*) is of the form:

$$i = I\,e^{-\varepsilon t}.$$

The time constant ε depends on the magnitude of C and on the neon tubes connected.

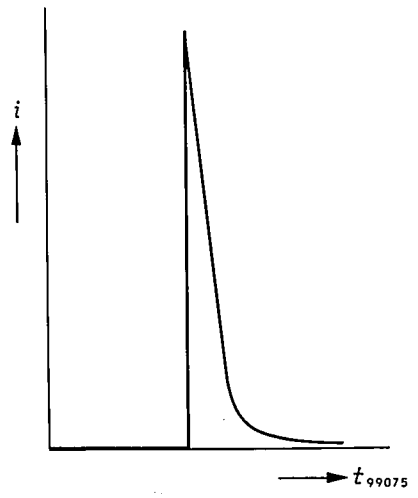


Fig. 10. Waveform of the discharge current i of the stray capacitance through the neon tube, on reignition.

The first discharge surge is followed by a series of current surges in increasingly rapid succession; their amplitude decreases whilst the instantaneous value i_s of the steady-state current increases. The explanation of this phenomenon is as follows. The first current surge, starting at $t = 0$ (*fig. 11*), leaves behind in the gas a strong concentration of free charge carriers, which can only disappear by recombination. This concentration, and hence the conductivity of the gas, is for a time so great

that the potential gradient of the discharge is virtually zero. The neon tubes then constitute (during an interval t_1-t_2) a short-circuit across the transformer. During this interval, the current through the tubes is for practical purposes equal to the short-circuit current i_k of the transformer secondary (the time constant ε of the current peaks is small compared with the time constant of the short-circuited secondary).

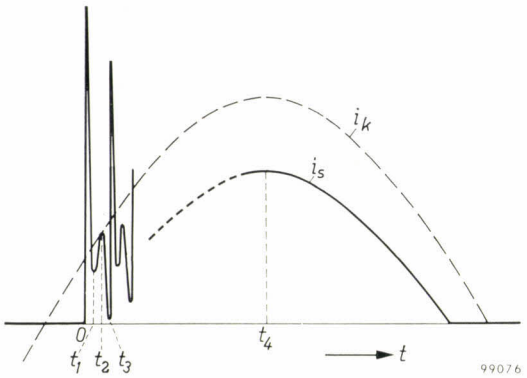


Fig. 11. Waveform of steady-state current i_s through a neon tube, showing superposition of discharge surges. i_k short-circuit current of transformer.

Thereafter the charge-carrier concentration is reduced by recombination. The potential gradient in the gas discharge therefore increases fairly rapidly, so that at the same time the current decreases somewhat (at t_2). The increase in the potential gradient is associated with the charging of the stray parallel capacitance C, but this can only take place at a rate determined by the damping factor α. The effect of this is that the voltage across C cannot increase fast enough, so that the discharge current must decrease still further (beyond t_2) until the discharge extinguishes (at t_3). Not until the potential across C is high enough does reignition take place, whereupon the process is repeated; see the oscillograms in *fig. 12*.

With each repetition the conditions gradually change, because up to the moment t_4 (fig. 11) i_s is still increasing. As a result, the variations in the potential gradient will gradually become smaller and the reignition voltage lower. Moreover, the current required to rapidly charge the stray capacitance becomes relatively less significant. This explains why the current surges become weaker as i_s increases, and follow each other more rapidly; this is to be seen in fig. 12. Provided the stray capacitance is not unduly high, the gas discharge will gradually approach the steady state. However, if C is so high that this state has not yet been reached at $t = t_4$, then, with i_s now decreasing, the current surges will gradually become stronger again and this can cause premature extinction of the gas discharge. This phenomenon is irregular in occurrence, causing the neon tube to burn unsteadily. The voltage oscillogram in fig. 12 shows two waveforms superimposed on one another, which is a consequence of the irregular extinction. The voltage pulse occurring upon the premature extinction is not always of sufficient amplitude to cause reignition. In that case the reignition of the tubes is also irregular. This phenomenon accounts for the upward trend of the right-hand part of the measured curve in fig. 9.

Effect of current surges on the life of neon tubes

During the first current surge an energy of $\frac{1}{2}CU_i^2$ is dissipated, U_i being the reignition voltage. In a

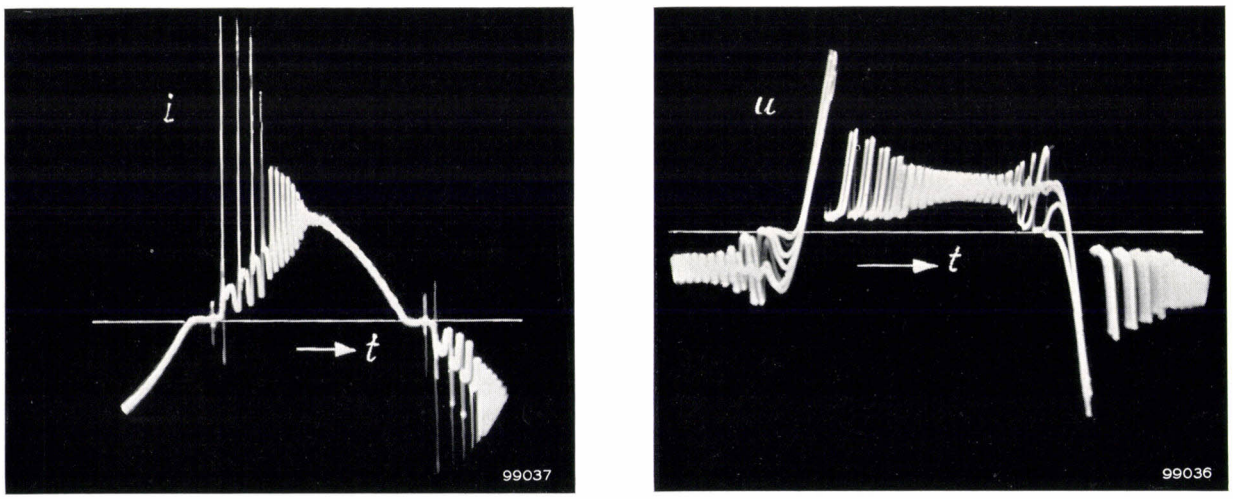


Fig. 12. Oscillograms of the current i through a neon tube and the voltage u across the tube. Owing to irregular extinction, two slightly differing voltage waveforms are seen superimposed in the photograph of u.

practical case (transformer 2×4 kV, 25 mA, which feeds a 6 m long neon tube through a balanced cable of 2×8 m), $C = 500$ pF and $U_i = 12.5$ kV. The energy of the surge is then 0.04 joule. The time in which this energy is dissipated is of the order of 1 microsecond, so that the power of the surge is of the order of 40 kW. Surges of this magnitude have a number of harmful consequences:

1) The tube electrodes are exposed to severe sputtering, which is the main cause of neon being absorbed on the glass wall.
2) The electrodes enter into mechanical vibration (which manifests itself in a high-pitched hum); in the long run this is likely to cause leakage at the glass seals.
3) The surges are a source of radio interference.

We shall deal here only with the effect mentioned under (1); the fact that unduly heavy current surges shorten the life of a neon tube is the primary reason for taking measures to avoid them.

The sputtering process takes place as follows. Particles are dislodged from the electrodes and strike the surrounding glass wall. They entrain a number of gas atoms, and these are absorbed into the layer of metal that forms on the glass. The amount of neon that disappears in this way with each surge may be assumed to be proportional to the average number of metal particles dislodged per surge and proportional to the number of neon atoms N per unit volume. The dislodging of the particles being a process involving an activation energy, the average number of particles dislodged per surge can be put as proportional to the energy P of the current surges per unit time, giving:

$$-V\,\mathrm{d}N = a\,N\,P\,\mathrm{d}t,$$

where a is a proportionality factor. If N_0 is the original number of neon atoms per unit volume, we obtain:

$$\ln\frac{N_0}{N} = a\frac{P}{V}t.$$

The neon tube has reached the end of its useful life ($t = T$) when N has dropped to a value N_T:

$$\ln\frac{N_0}{N_T} = a\frac{P}{V}T,$$

whence:

$$T = \left(\frac{1}{a}\ln\frac{N_0}{N_T}\right)\frac{V}{P} = A\frac{V}{P}, \quad \ldots \quad (10)$$

where the constant $(1/a)\ln(N_0/N_T)$ is written as A.

On the assumptions made above, the life should be inversely proportional to the energy P of the surges per second, and, if P is constant, directly proportional to the gas volume V; thus, for a given tube diameter and initial pressure, the life is directly proportional to the tube length l_b.

This explains the fact, known from experience, that if two neon tubes of different length are connected in series, the long tube will have a longer life than the short one (the power P is dissipated equally over the four electrodes, at least at the beginning). Series-connected tubes should therefore differ as little as possible in length. This restriction apart, no useful purpose is served by dividing the total tube length into sections. This does not appreciably reduce the reignition voltage, nor therefore the total energy of the current surges. Since a certain loss of energy occurs in the gas discharge at each electrode, the total loss is increased by the greater number of electrodes, and the total tube length must therefore be smaller than in the case of one long tube. A smaller length per tube means a smaller volume of gas, and therefore, according to (10), a shorter life.

The conclusion just drawn, viz. that $T \propto l_b$, is valid for $P =$ constant, which is the case in the comparison of series-connected neon tubes. We now compare two installations consisting of the same number of neon tubes connected in series, but where the tubes in the one installation are n times longer than in the other. In each of the installations the transformer voltage must be matched to the overall tube length, in which case (where the capacitance C is constant) the life T is *inversely* proportional to l_b: since the required reignition voltage is approximately proportional to l_b, and P is thus proportional to l_b^2, it follows from (10) that $T \propto l_b^{-1}$.

If the high-tension cable is fairly long, the transformer capacitance is negligible compared with the cable capacitance C_k, which is proportional to the length l_k of the cable. The power P is then approximately proportional to l_k, so that, where d is the diameter of the neon tubes, it follows from (10) that

$$T \propto \frac{d^2}{l_b l_k}. \quad \ldots \ldots \quad (11)$$

In the Philips factory at Roosendaal, The Netherlands, life tests have been carried out on neon tubes of different lengths and diameters, and connected by cables of varying lengths. Some results are given in *fig. 13*. Fig. 13*a* shows the average life T as a function of cable length l_k, for tubes of various lengths l_b. Fig. 13*b* shows T as a function of the neon volume V, for various cable lengths l_k. The neon

tubes were connected in series (P therefore being constant). The results are in broad agreement with the formulae (10) and (11). The fact that fig. 13*a* shows the life to be roughly inversely proportional to l_k, and therefore inversely proportional to the cable capacitance, supports the premise assumed earlier that the average number of metal particles dislodged per surge is proportional to the energy P of the surges per unit time.

From (10) and (11) one would expect to find in fig. 13*a* and *b*, in which both scales are logarithmic, straight lines having a slope of exactly −45° and +45°, respectively. The fact that the slopes actually found differ somewhat from 45° is probably attributable to the fact that, owing to the series-connection of tubes of different lengths, the gas disappeared most rapidly from the shortest tube; the shorter the tube, the more rapidly the gas pressure declined. After some time the pressures will have become different and the energy of the current surges will no longer have been uniformly distributed over the electrodes.

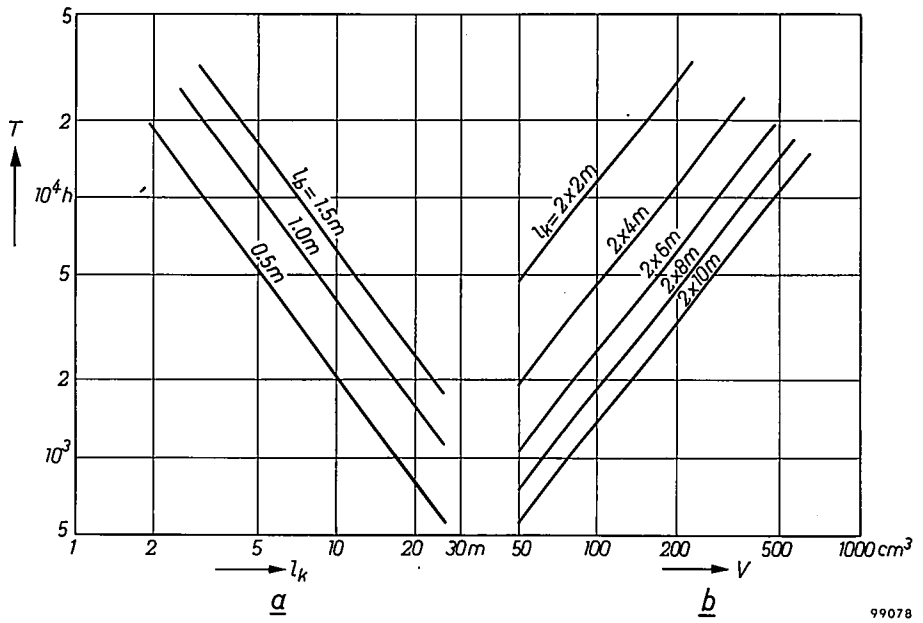


Fig. 13. Results of life tests on neon tubes carried out in the Philips factory at Roosendaal. *a*) Average life T as a function of cable length l_k, for tubes of various lengths l_b. *b*) T as a function of gas volume V for series-connected tubes, for various cable lengths l_k. All tubes used for these tests had the same cross-section.

According to (11) it is possible with shorter tubes (and appropriately smaller transformers) to obtain a longer life or, for the same life, one can use a longer cable. This is illustrated by some examples in *fig. 14a-d*. Assuming that the distance AB from the transformer to the one end of the neon tubes is the same in these four examples, it can be seen with the aid of the above considerations that cases *a* and *b* are completely equivalent, both as regards life and cable length. In case *c* we may expect the life to be about 1.5 times longer, the cable length required being also 1.5× longer. In case *d* the life is about twice as long, and the cable length also twice as long. The voltage on the cable is much lower for *c* and *d* than for *a* and *b*.

Surge limitation

The foregoing considerations have shown clearly the desirability of finding some means of limiting the current surges and of dissipating the energy of the charged stray capacitances as far as possible *outside* the neon tubes.

The surges can be limited by means of a resistor or a choke. Being cheaper, resistors are frequently used for this purpose. The method is to connect the resistor to the end of the cable in series with the neon tubes. The effect is slight, however, since the value of the resistance R_1 must be restricted to about 10% of the load resistance represented by the neon tubes; otherwise the loss in the resistance would demand too great a sacrifice in the length of tube in relation to the transformer voltage. The current peaks are then reduced by a factor of about 2, which is hardly adequate.

If the current surges are limited by a choke (inductance L_1, resistance R_1) at the end of the cable, the current surge is given approximately by:

$$i = \frac{E}{L_1} \exp\left(-\frac{R_1}{2L_1}\, t\right) \frac{\sin qt}{q},$$

where q is given by

$$q = \sqrt{\frac{1}{L_1 C} - \frac{R_1^2}{4L_1^2}}.$$

q may be either real or imaginary. The special case, $q = 0$, leads to:

$$i = \frac{E}{L_1} t \exp\left(-\frac{R_1}{2L_1} t\right).$$

The peak value of i is then $2E/eR_1 = 0.74\ E/R_1$, and thus little less than in the case where only a resistance R_1 is used. Imaginary values of q are therefore useless here. In order to examine the possibilities presented by the use of real values of q (i.e. periodic solutions), we consider the case

$$\frac{1}{L_1 C} \gg \frac{R_1^2}{4L_1^2}.$$

The peak value I of the current i is then

$$I \approx E \sqrt{\frac{C}{L_1}}.$$

It is evident that in this way a fairly small inductance L_1 (provided its own capacitance is small enough) is able to reduce the current surges by a factor of, say, 10. The condition is, however, that the oscillations produced in the L_1-C circuit should be sufficiently damped. This is necessary, 1) in order to dissipate most of the surge energy outside the neon tubes, and 2) because otherwise the tubes would again burn unsteadily, for the same reason as described above.

The oscillations can be damped either by shunting a resistance across the coil, or by ensuring that the iron losses in the core of the coil are sufficiently high.

Although the unfavourable effects of high cable capacitances can be appreciably reduced in this way, when long cables are necessary it is advisable in addition to use low-capacitance cables. For example, the conventional lead-sheathed cable, which has a capacitance of 120 pF per metre, can advantageously be replaced by a rubber-insulated wire enclosed in 5/8″ piping, which has a capacitance of only 65 to 70 pF per metre.

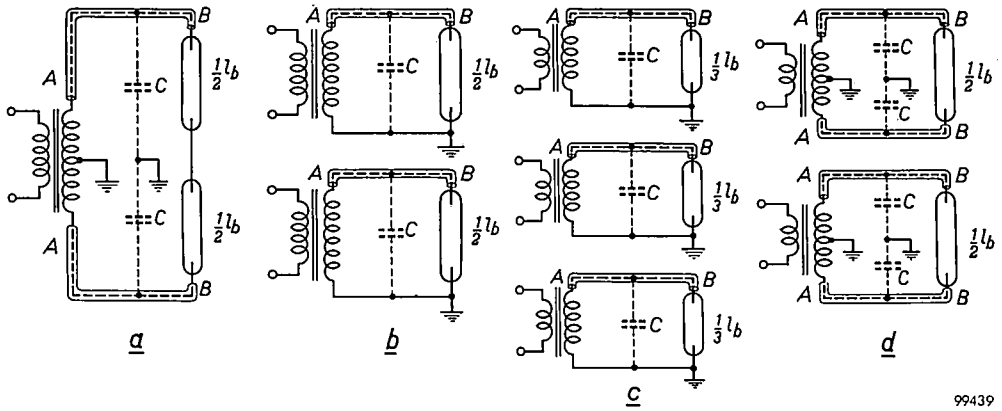


Fig. 14. Various circuits for the supply of neon tubes. The total length l_b of the tubes is the same in all four cases. Similarly, the distance AB is assumed to be the same in all four cases.
a) Two tubes, each $\frac{1}{2}l_b$ in length, fed by a balanced transformer.
b) Each tube fed by its own transformer. Equivalent to (*a*).
c) Three tubes, each $\frac{1}{3}l_b$ in length, each fed by its own transformer. The life is about 1.5 × that in (*a*) and (*b*). The required length of cable is also about 1.5 × that of (*a*) and (*b*).
d) Two tubes, each $\frac{1}{2}l_b$ in length, each fed by its own balanced transformer. Here the life is about 2 × that of (*a*) and (*b*), and the required length of cable is also twice as long.

Unbalanced cables

Where unbalanced transformers are used (one side of the secondary being earthed, see figs. 2*a* and *c* and 14*b* and *c*), the effect of cable capacitance can be limited by connecting the unearthed secondary terminal by as short a lead as possible to the neon tube; the other (earthed) lead may be as long as required.

In the case of balanced transformers (where the centre of the secondary winding is earthed, see fig. 14*a* and *d*) it is possible to limit the current surges by unbalanced connection of the cables, that is by making one high-tension cable much shorter than the other. During the surge the two stray capacitances are effectively in series for the discharge; if the one capacitance is much smaller than the other, the equivalent capacitance is virtually equal to the smaller capacitance, which therefore mainly determines the strength of the surges. One might thus be inclined to make the short cable extremely

short, but in that case, if the other connection is long (> 10 m), complex transients may arise after the reignition, and these will almost immediately extinguish the discharge. However, with a length ratio of 1 : 10 this effect is not troublesome, and moreover the life of the neon tubes is appreciably lengthened in this way without other measures being necessary.

Finally, some remarks on the maximum voltage appearing across the cables. When the installation is operating, this voltage in the case of unbalanced transformers is equal to the reignition voltage, and in the case of balanced transformers it is equal to half the reignition voltage. In practice this amounts in both cases to a peak voltage of at the most 6.5 kV. When the transformer is not loaded (neon tubes not burning) the maximum voltage on the cable is $1/(1-\omega^2 LC)$ times higher than with an infinitely short cable. Where long cables are used, this means an increase of 10 to 15% above the no-load voltage that would be present with an infinitely short cable.

Summary. The high voltage required for the reignition of neon tubes would call for relatively expensive power transformers if use were not made for this purpose of the transient produced periodically when the current has dropped to zero. This transient consists of a damped oscillation of a frequency (500-1200 c/s) partly determined by stray capacitances. The height of the first positive voltage peak governs the length of tube that can be connected to a given transformer; it is affected by the quality of the laminations of the transformer core, by the magnitude of the magnetic induction in the core, and to a certain extent by the capacitance (hence the length) of the high-tension cable(s) between transformer and neon tube.

With each periodic reignition of the neon tubes the stray capacitances discharge through the neon tube. These strong current surges cause the tubes to burn unsteadily and shorten their useful life. The surges can be limited by constructing the installation from a number of smaller units, by using unbalanced high-tension cables of low capacitance, and by connecting a small inductance in series with the neon tubes.

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 Laboratories, Eindhoven, Netherlands.

2681: J. Halberstadt: Some experiments with radioactive preparations of 2,4,5,4′-tetrachlorodiphenylsulphone, a new acaricide (Meded. Landbouwhogesch. Opzoekingsstat. Gent **23**, 788-794, 1958, No. 3/4).

Radioactive-isotope tracer technique has been used to study the metabolism of "Tedion V18" (an acaricide against mites) in plants and animals. The active part of Tedion, viz. 2,4,5,4′-tetrachlorodiphenylsulphone, was labelled by partly replacing the sulphur with the isotope ^{35}S. Apple trees were sprayed with the labelled preparation, formulated as a wettable powder and as a miscible oil. From measurements of the radioactivity, it was found that an active residue remains on the leaves for a long time; the Tedion that permeates in the leaf is depleted by chemical change and transport but is replaced by constant permeation from outside. Miscible oil appears to be a more suitable formulation than wettable powder. Administered to rats (a maximum of 100 mg per kg body weight), 40-45% of a Tedion dose is found unchanged after 48 hours in the faeces; the remainder is broken down.

The substance of this paper will shortly appear in the Philips Technical Review.

2682: W. Duyfjes: Some problems in pesticide formulation (Meded. Landbouwhogesch. Opzoekingsstat. Gent **23**, 837-845, 1958, No. 3/4).

Discussion on the formulation of pesticides, i.e. chemicals used against insects, mites, fungi and weeds. Particular attention is paid here to pesticides prepared for spraying with water. It is shown that for each pesticide a separate investigation is necessary to determine the best formulation. Not only must the physical and chemical properties be considered (e.g. the stability of suspensions or emulsions), but the formulation must be tested for its biological effectiveness. Sometimes formulations with the best physico-chemical characteristics give less favourable biological results, and vice versa. Examples are given by way of illustration. See also Philips tech. Rev. **19**, 165-176, 1957/58.

2683: J. H. Stuy: Nucleic acid synthesis in ultraviolet-irradiated Bacillus cereus (J. Bacteriol. **76**, 668-669, 1958, No. 6).

Further investigations into the function of nucleic acids in physiological processes in bacteria (see also No. **2635** of these Abstracts). Kelner has shown that U.V. irradiation inhibits the

formation of deoxyribonucleic acid (DNA) by the organism. Some time after stopping the irradiation, the formation of DNA can once more begin. This paper concerns an investigation into the composition of DNA and RNA (ribonucleic acid) as these are formed after irradiation in *Bacillus cereus* p2. With the somewhat crude chemical methods used, no difference could be detected between the DNA and the RNA formed before the irradiation. To what extent the post-irradiation DNA differs biologically from that normally formed, is a problem that will be further investigated.

A 6: E. Baronetzky: An apparatus, capable of being heated, for high-vacuum grinding of solids (J. sci. Instr. **35**, 427-428, 1958, No.11).

A rotary file is mounted on one end of a spindle free to rotate in glass-metal bearings. At the other end of the spindle is a steel rotor with a layer of copper brazed to its periphery. This whole system together with the material to be ground and the feed guides are surrounded by an envelope of hard glass and can be heated to 400 °C. The drive is via a multi-poled magnetic ring of ferroxdure, concentric with the rotor and outside the glass envelope, itself driven by a variable speed DC motor.

A 7: A. Rabenau and P. Eckerlin: $BeSiN_2$, eine neue Verbindung mit Wurtzitstruktur (Naturwiss. **46**, 106-107, 1959, No. 3). ($BeSiN_2$, a new compound having a wurtzite structure; in German.)

Note concerning a new compound $BeSiN_2$, one of the group of $A^{II}B^{IV}X^{V}_2$. It has the wurtzite structure analogous to the nitrides of Al, Ga and In, and is colourless. The crystallographic data are: $a = 2.872 \pm 0.004$ Å, $c = 4.674 \pm 0.004$ Å. The X-ray density was found to be 3.24 g/cm³; the value from weighing and volume determination was 3.12 g/cm³.

A 8: H. G. Grimmeiss and H. Koelmans: Über die Kantenemission und andere Emissionen des GaN (Z. Naturf. **14a**, 264-271, 1959, No. 3). (On edge emission and other kinds of emission of GaN; in German.)

The fluorescent properties of pure GaN and GaN doped with Li, Zn and Mg have been measured under U.V. radiation and cathode rays as functions of the temperature. Several emission bands have been found. The shortest-wavelength band is attributed to recombination of free electrons with free holes.

A 9: F. Kettel: Die Wärmeleitfähigkeit von Germanium bei hohen Temperaturen (Phys. Chem. Solids **10**, 52-58, 1959, No. 1). (The thermal conductivity of germanium at high temperatures; in German.)

An account is given of thermal conductivity measurements on germanium single crystals, both intrinsic and with high hole conductivity. The author took mobility values, and values for the width of the forbidden zone and its dependence on temperature — all necessary for theoretical discussion of the electric component of thermal conductivity — from the measurements of other authors, and confirmed them by comparison with his own measurements of thermoelectric force and electrical conductivity. The measured values of the electronic component of thermal conductivity, of both intrinsic and hole-conductive germanium, are about 4 times as high as those determined theoretically.

A 10: A. Goertz: Zur Theorie der diffusen Reflexion und Transmission beim Vorliegen elastischer Vielfachstreuung (Z. Physik **155**, 263-274, 1959, No. 3). (On the theory of diffuse reflectivity and transmittivity in case of elastic multiple scattering; in German.)

In order to determine the diffuse reflectivity and transmissivity of multiply scattered particles, there is not necessarily the need of solving the complete differentio-integral equation of multiple scattering. It can be shown that the problem reduces to solving a set of two simultaneous differential equations of the first order, provided that the amount of sideward scattering may be neglected. The solutions of these equations furnish all data required, if the mechanism of single scattering is known.

A 11: E. Kauer, O. E. Klinger and A. Rabenau: Über Leitfähigkeitsmessungen im System ZrO_2-MgO (Z. Elektrochemie **63**, 927-936, 1959, No. 8). (Electrical-conductivity measurements on the system ZrO_2-MgO; in German.)

The electrical conductivity of pure ZrO_2 and of the system ZrO_2-MgO has been measured as a function of the temperature between 500 and 1500 °C. Further, thermoelectric measurements have been carried out, which indicate positive charge carriers. The measurements indicate that the conductivity (at least at the partial pressure of oxygen in the air) is preferentially due to oxygen vacancies. It is shown that, with a knowledge of the conduction mechanism, conductivity measurements as a function of time and temperature give a sensitive method of following the solid-state reaction involved. It not only gives a picture of the phase concentrations but also gives some idea of the progress of the transformation with time.

R 376: K. Jost: Impedance transformations through lossless two-ports represented by fractional linear transformations of the unit circle (Philips Res. Repts. **14**, 301-326, 1959, No. 4).

Any impedance transformation by a lossless two-port network corresponds to an automorphism of the unit circle in the complex reflection-coefficient plane. Such a mapping of the unit circle, determined by three parameters, can always be resolved into three successive elementary transformations: a rotation about the origin, a hyperbolic transformation corresponding to the impedance transformation by an ideal transformer, and a further rotation about the origin. Resolution in this way leads to a simple graphical method for the determination of the transformed impedance and for the treatment of cascades of lossless two-ports. Cascades giving an ideal transformer are investigated. The determination of the parameters of a two-port (calibration of the two-port) is described.

R 377: M. T. Vlaardingerbroek: Noise in electron beams and in four-terminal networks (Philips Res. Repts. **14**, 327-336, 1959, No. 4).

It is shown that the calculation of the minimum noise figure of electron-beam amplifiers, obtainable by varying the physical properties of the lossless beam-transducer between the electron gun and the interaction region, can be performed in a manner which is similar to the calculation of the minimum noise figure of network fourpoles based on the method of varying the signal-source impedance. Further, the similarity between the propagation of noise fluctuations along electron beams and the transformation of noise sources across passive fourpoles is emphasized.

R 378: P. Penning: Rate of diffusion-limited annihilation of excess vacancies (Philips Res. Repts. **14**, 337-345, 1959, No. 4).

The rate of removal and the spatial distribution of excess vacancies are calculated for the case where the transport to vacancy sinks takes place by diffusion and where dislocations in the volume of the sample, the surface of the sample, or both may act as sinks. The medium in which the diffusion takes place is considered as a continuum. It is found that in the case where the dislocations are the only sinks, the decay in average concentration is exponential. The decay time constant is almost inversely proportional to the dislocation density. In the case where the surface is the only sink an appreciable deviation from an exponential decrease in average concentration does occur in the beginning of the removal process. In a bar of square cross-section the contribution of the two types of sink is about equal if a few tens of dislocations emerge through the smallest cross-section.

R 379: H. U. Harten: Surface recombination of silicon (Philips Res. Repts. **14**, 346-360, 1959, No. 4).

The surface-recombination velocity of electrons and holes in silicon is investigated by measuring the photovoltaic effect of a p-n-junction alloyed on a thin silicon wafer. The method has been reported previously; its principle is explained here on the basis of a hypothetical experiment. It appears from the measurements that after treating the surface with an aqueous solution of potassium dichromate the recombination is lowered by ozone in the ambient atmosphere and raised by moisture. The opposite behaviour is observed after etching with hydrofluoric acid. In principle there is no difference in the behaviour of n-type and p-type silicon. In many cases the surface recombination can be decreased with light, particularly if the surface treatment tends to form an inversion layer. The observations can be interpreted by the assumption that the surface recombination on silicon is due to recombination centres and therefore influenced by the position of the Fermi level at the surface, and that an additional influence is due to the voltage across the surface barrier layer.

R 380: D. de Nobel: Phase equilibria and semiconducting properties of cadmium telluride (Philips Res. Repts. **14**, 361-399, 1959, No. 4; continued in No. **R 384**).

In this thesis (Leyden, May 1958) the relation is studied between the electrical and optical properties of single crystals of cadmium telluride and the conditions of preparation. The p-T-x diagram of the system cadmium-tellurium is described, showing the temperatures and Cd pressures at the maximum melting point and at the melting point of stoichiometric CdTe. The compound is purified by zone refining; foreign atoms are incorporated by zone levelling and single crystals are grown, which are reheated at various Cd pressures between 700 and 1000 °C and then quenched to room temperature. On these samples conductivity and Hall measurements are performed at various temperatures, which lead to values of the concentration of charge carriers, of the ionization energies of the various centres and — for n-type samples with shallow donors — of the density-of-state effective mass of the free electrons ($m_n^*/m = 0.14 \pm 0.04$). Thermoelectromotive-force measurements of various n and p-type samples

lead to values of the effective mass for both types of carrier, depending on the value adopted for the transport energy of electrons and holes (m_n^*/m = 0.13-0.066; m_p^*/m = 0.41-0.22). From a hydrogen-like model for the shallow donor, an inertial effective mass for the electrons of $(m_n^*)_i = 0.147$ is obtained. From optical transmission measurements at various temperatures the band gap is found to be 1.50 eV at room temperature and the temperature dependence 2.34×10^{-4} to 5.44×10^{-4} eV/°K. Peaks in the spectra of photoluminescence and photoconductivity can be correlated with a band-band transition and with transitions between levels caused by known centres, and one of the bands. A theoretical discussion is presented of the various equilibria which determine the state of CdTe at high temperatures. By assuming a certain band scheme and certain values for the equilibrium constants, it is possible to calculate the concentrations of charge carriers and centres at room temperature as a function of the Cd pressure at which the crystals were reheated. A qualitative comparison between the experimental and various theoretical diagrams leads to the adoption of a definite band scheme, in which two levels are attributed to the cadmium vacancy and one each to interstitial cadmium, to indium (as a specific donor) and to gold (as a specific acceptor). From a quantitative discussion of the diagrams the values of some equilibrium constants are obtained, viz. the Frenkel constant K_F and the reduction constant K_r, describing the equilibrium: crystal-vapour. The temperature dependence of these constants leads to the activation energies required for the atomic processes involved. Finally, the association effects which take place during the quenching of activated samples are analysed. In the appendix a discussion is given of the type of bonding in CdTe.

R 381: Joshua Ladell and William Parrish: Determination of spectral contamination of X-ray tubes (Philips Res. Repts. **14**, 401-420, 1959, No. 5).

An X-ray method is outlined for the qualitative and quantitative analysis of the spectral purity of X-ray tubes. The method, based on well-known principles, employs a standard diffractometer equipped with a xenon-filled proportional counter and molybdenum-foil analyser. The various theoretical and practical aspects of the determination of the correction factors for the comparison of different wavelengths are described.

R 382: S. Duinker: Generalization to non-linear networks of a theorem due to Heaviside (Philips Res. Repts. **14**, 421-426, 1959, No. 5).

A theorem enunciated by Heaviside and proved by Lorentz for linear electromagnetic systems subjected to a constant electric force which is suddenly impressed, is extended to electrical networks comprising non-linear reactances and linear dissipances. The theorem states that the amount of work to be done by a constant-voltage source exclusively to sustain the transients arising as the network changes from the rest state to the steady state, is equal to the excess of the sum of electric energy and co-energy over the sum of magnetic energy and co-energy.

R 383: J. F. Marchand and A. Venema: Note on superconducting tantalum films (Philips Res. Repts. **14**, 427-429, 1959, No. 5).

Thin superconductive tantalum films are difficult to produce. In this note it is shown how such films can be made by evaporation under extremely low pressure. The films thus prepared have been found to exhibit practically the same superconducting properties as the bulk material.

R 384: D. de Nobel: Phase equilibria and semiconducting properties of cadmium telluride (Philips Res. Repts. **14**, 430-492, 1959, No. 5). Continuation of **R 380**.

Now available:

R. van der Veen and G. Meijer: Light and plant growth (Philips Technical Library 1959, pp. 164, 92 illustrations).

The first five chapters of this book are devoted to the various processes in plants that are caused or influenced by light; the sixth chapter deals with the practical application of this knowledge, in particular with regard to the irradiation of plants by artificial light. The titles of the chapters are: I. Light measurement for plant irradiation, II. Photosynthesis, III. Phototropism, phototaxis and photonasty, IV. Photoperiodism, V. Effect of the colour of the light, VI. The use of artificial light in horticulture.

VOLUME 21, 1959/60, No. 8 pp. 221-252 Published 30th June 1960

Philips Technical Review

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

REFLEX KLYSTRONS FOR WAVELENGTHS OF 4 AND 2.5 mm

by B. B. van IPEREN.

621.385.6.029.65

In recent years there have been attempts on every hand to produce shorter and shorter wavelengths, both with the familiar types of microwave tubes and also with devices operating on different principles. In this respect the reflex klystron has proved to be capable of a surprisingly good performance. A few years ago Philips were the first to bring out a reflex klystron for 4 mm waves, and now good results have also been achieved with a klystron for 2.5 mm waves.

Reflex klystrons for millimetre wavelengths are mainly used in radar receivers, in waveguide communication systems, and for various research purposes, e.g. for microwave spectroscopy and temperature measurements on gas-discharge plasma (from the propagation properties of microwaves in a plasma one can derive the electron density and temperature of the plasma). The latter application is of especial importance in that it allows the determination of the enormously high gas temperatures, of the order of millions of °K, required to initiate thermonuclear reactions.

The reflex klystron has proved to be an eminently useful type of tube for generating microwaves. Its power output is of the order of 10 to 1000 mW, which is quite adequate for many purposes. Articles on a reflex klystron for 4 mm waves appeared in this journal and elsewhere some years ago [1]). Tubes of this type are now being manufactured by Philips in small numbers and with reasonable reproducibility [2]). Some details of these are given below. We then deal with theoretical considerations concerning the production of even shorter waves, and which have led to the design of a reflex klystron for operation at about 2.5 mm wavelength. This tube, with which a wavelength as short as 2.3 mm has been achieved, and which is still in the experimental stage, will be discussed at the end of this article.

In its simplest form a reflex klystron consists of an electron gun, a resonant cavity and a repeller electrode (*fig. 1*). The electron gun "shoots" a beam of electrons through a resonant cavity which contains two central holes and also constitutes the anode. After passing through the resonant cavity, where they are modulated in velocity (by the minute oscillation already present, due to noise fluctuations or the transient resulting from switching-on the

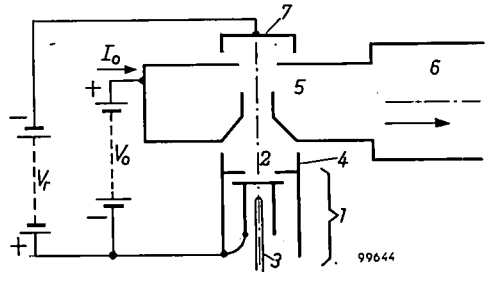


Fig. 1. Schematic axial cross-section of a reflex klystron. *1* electron gun with cathode *2*, heater *3* and focusing electrode *4*. *5* resonant cavity (also anode) with output waveguide *6*. *7* repeller electrode. V_0 accelerating voltage. I_0 beam current. V_r repeller voltage.

electron beam), they are returned by the repeller, which has a negative potential V_r with respect to the cathode. Provided V_r and the distance from the repeller to the cavity are properly chosen, the returning electrons again pass through the cavity in such a phase that, by interaction with the electromagnetic resonator field, they give up energy to the resonant cavity. The oscillation is thus

[1]) B. B. van Iperen, A reflex klystron for 4 mm waves, Philips tech. Rev. **18**, 51-52, 1956/57; Klystrons reflex pour ondes millimétriques, Le Vide **11**, 264-266, 1956.

[2]) The last stage in the development of this tube was the work of H. W. van der Voorn, now in charge of its manufacture.

amplified, so that subsequent electrons are more strongly modulated in velocity and so on. The system thus acts as a generator.

If the tube is to operate at an accelerating potential V_0 lower than e.g. 1000 V, more elaborate constructions are necessary, making use of grids in the resonant cavity, or of a ribbon-shaped or hollow electron beam. Where the aim, however, is to generate oscillations of the highest possible frequency, grids can better be dispensed with, since electron bombardment overheats them and therefore limits the maximum power per unit surface area. One may therefore expect tubes without grids to be capable of generating higher frequencies, and our investigations have accordingly been limited to this particular type. Our object being to develop tubes for continuous operation, attention had to be paid to the dissipation problem in spite of the absence of grids.

To simplify as far as possible the technological problems involved, a design with axial symmetry was adopted, using a "solid" electron beam — in principle, therefore, as illustrated schematically in fig. 1.

The DX 151 reflex klystron for 4 mm waves

The type DX 151 reflex klystron for 4 mm waves now being manufactured is a further development of the tube dealt with in the article mentioned under footnote [1]). *Fig. 2* represents a cross-section of the central portion, showing the electron gun with its components *2*, *3* and *4*, the resonant cavity *5*, the output waveguide *6* and the repeller electrode *7*. Diametrically opposite the output waveguide is a second waveguide *8*, containing a shorting plunger for adjusting the optimum load impedance. Mica windows in the waveguide provide the necessary vacuum seal. The output waveguide is coupled to the resonant cavity via a matching transformer; this consists of a gap *9* measuring about $\frac{1}{4}$ wavelength in the axial direction, an annular space *10* and the waveguide *8* with shorting plunger. One wall of the space *10* is so thin at one point (at *11*) as to

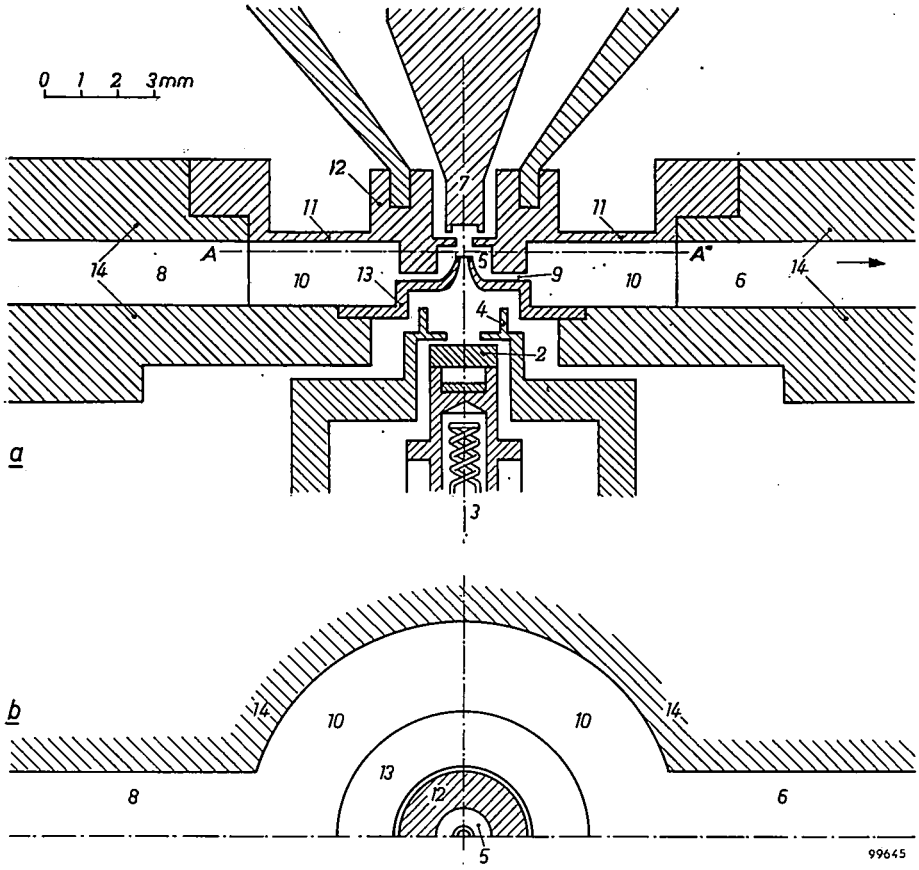


Fig. 2. Central portion of type DX 151 reflex klystron for 4 mm waves. *a*) axial section, *b*) transverse section along horizontal plane through *A-A* in (*a*). *2* L cathode. *3* heater. *4* focusing electrode. *5* resonant cavity. *6* output waveguide. *7* repeller. *8* waveguide with shorting plunger (not shown). *9* quarter-wave transformer. *10* annular space (*9*, *10* and *8* together form the matching transformer between the resonant cavity *5* and the output waveguide *6*). *11* thin wall allowing copper section *12* to be moved axially in relation to the copper section *13* by means of an external tuning mechanism (mechanical tuning). *14* copper block in which central portion is mounted. Dimensions:

Diameter of opening in electrode *4*	1.0 mm	Diameter of holes in *12* and *13* . .	0.25 mm
Diameter of resonant cavity *5* . .	1.6 mm	Width of waveguides *6* and *8* . .	3.6 mm
Height of resonant cavity *5* . . .	0.7 mm	Height of waveguides *6* and *8* . .	1.8 mm

enable the upper part of the tube to be moved axially in relation to the bottom part by means of an external tuning mechanism. This displacement varies the gap width of the interaction space and hence the equivalent capacitance of the resonant cavity. In this way the resonance frequency can be varied continuously over a range of about 10 to 15%, for a displacement of about 100 μ.

Because of its complexity is not possible to calculate directly the properties of the matching transformer. However, its mirror symmetry allows a simple equivalent circuit to be given, from which insight can be obtained into the operation of the shorting plunger as a matching element. In this circuit, shown in *fig. 3*, the resonant cavity is replaced by an *L-C-R* network. This is connected by a lecher line (electrical length x and characteristic impedance $2Z_0$) to two other lecher lines in series (characteristic impedance Z_0), one of which is terminated by the load (ordinarily with impedance Z_0) and the other by a sliding shorting bridge corresponding to the plunger. When this bridge is positioned such that the distance to the branching point is $2n+1)\lambda/4$, where n is a whole number, the load conductance measured across the circuit is zero, indicating that the load is entirely decoupled from the resonant cavity. This means that the load can be reduced to any desired value by moving the bridge. In tubes for mm waves this is a very desirable feature, particularly where the object is to achieve the shortest possible wavelength.

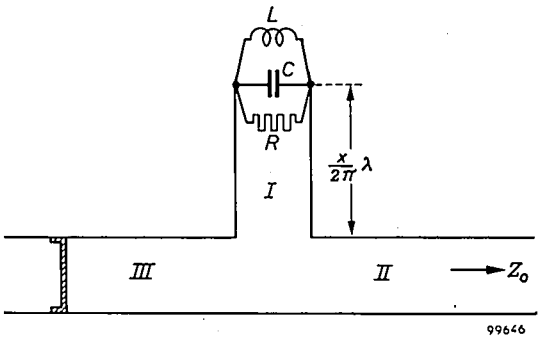


Fig. 3. Equivalent circuit of matching transformer *9-10-8* in fig. 2. *L-C-R* represents the resonant cavity *5*. *I* is a lecher line with characteristic impedance $2Z_0$ and electrical length x; *II* and *III* are lecher lines of characteristic impedance Z_0, replacing waveguides *6* and *8*, respectively.

The maximum load conductance obtainable depends on x and Z_0. In practice the tendency is to design the matching transformer in such a way that x and Z_0 in the required frequency band have values that make it possible to vary the load over a sufficiently wide range.

Fig. 2 also shows some constructional details. The waveguides *6* and *8*, the annular space *10* and the place for the copper sections *12* and *13*, which together constitute the resonant cavity and the $\frac{1}{4}$-wave transformer, are cut in a single block of copper *14*. Parts *12* and *13* are made by a hot-pressing process, the high precision of which benefits the reproducibility of the tube characteristics.

One of the most difficult technical problems in tubes of this kind is the accurate alignment of the electrodes; in the present case the alignment must be accurate to within 10 μ. This was solved satisfactorily by mounting the electron gun and the repeller in such a way that, after the tube is sealed off, they can still be moved fractionally in various directions perpendicular to the axis of the tube (after-alignment). For this purpose a flexible diaphragm is used as part of the tube wall. *Fig. 4* shows the arrangement in the case of the repeller; a similar arrangement is adopted for the electron gun.

For reasons explained below, a very high current density is required from the cathode in such tubes. This makes the use of a dispenser-type cathode, e.g. an L cathode [3]), essential. Moreover, the L cathode readily fulfils another condition, namely that the emitting surface must be smooth to within about 1 μ in order for the electron beam to have the correct shape. The L cathode used in the DX 151 tube delivers the required current density of 2.5 A/cm^2 for an average useful life of several thousands of hours.

The diameter of the electron beam at the narrowest point is about 0.12 mm, i.e. roughly half the diameter of the holes in the resonant cavity. A beam as narrow as this is required in order that the majority of the electrons will still pass through the resonant cavity in spite of the beam having widened out in its passage to and fro in the repeller space.

Fig. 5 shows the output power (the average over several tubes) as a function of frequency and wavelength. A photograph of the tube, complete with

[3]) H. J. Lemmens, M. J. Jansen and R. Loosjes, A new thermionic cathode for heavy loads, Philips tech. Rev. **11**, 341-350, 1949/50.

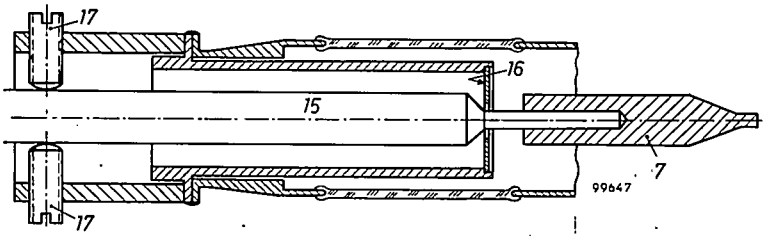


Fig. 4. The repeller electrode *7* is fitted to a stem *15*: this is fixed vacuum-tight in a flexible diaphragm *16*. By adjusting screws *17* the alignment of the repeller in relation to the axis of the stem can be altered whilst the tube is operating. Two other screws permit adjustment in a direction at right-angles to the first. The same system is used for after-alignment of the electron gun.

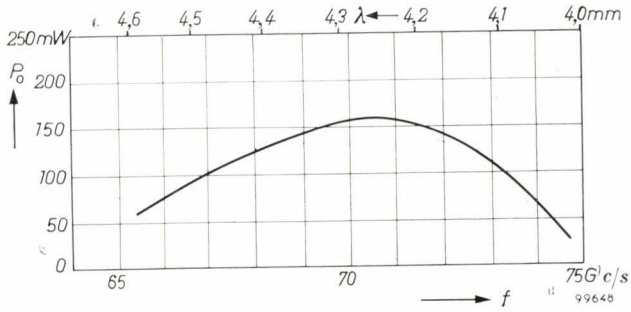


Fig. 5. Power output P_o of type DX 151 reflex klystron (average value over several tubes) as a function of frequency f (in Gc/s = 10^9 c/s) and wavelength λ, varied by the tuning mechanism. The repeller voltage was adjusted to the most favourable value for each measurement.

tuning mechanism and further accessories, can be seen in *fig. 6*: the principal electrical data are given in the caption.

Possibilities of generating even shorter waves

The results obtained with the 4 mm klystron prompted a theoretical investigation into the possibilities of generating even shorter waves with tubes of the same kind. In this connection it is useful first of all to compare the practical results with those derived from the simple theory of reflex klystrons. According to the elementary theory the efficiency, i.e. the proportion of the D.C. power input which is converted into useful high-frequency power, is given by the formula [5]):

$$\eta = \frac{k\beta_2}{2\pi N\beta_1} X^2 J_2(X). \quad \ldots \quad (1)$$

In this expression,

k is the fraction of the original electron beam that passes through the interaction space of the resonant cavity on the way back,

N is the transit time of a non-velocity-modulated electron in the repeller space, expressed in terms of the period of the high-frequency voltage (N is a whole number + $\frac{3}{4}$),

β_1 is the coupling factor of the forward beam with the resonant cavity, and

β_2 is the coupling factor of the returning beam with the cavity (for explanation see below),

X is the bunching parameter (this term will also be explained presently), and

J_2 is the Bessel function of the first kind and the second order.

The meaning of the coupling factors β_1 and β_2 is most readily understood in the case of velocity modulation by means of two grids between which there is an alternating voltage $\tilde{V} \cos 2\pi t/T$. An electron passing between this pair of grids in a time which is very small compared with the period T undergoes a change in kinetic energy equal to $e\tilde{V} \cos 2\pi t_M/T$, where t_M is the moment at which the electron is situated midway between the two grids. If the time taken to pass between the grids is not very small compared with T, a correction factor β must be applied. The value of β is independent of t_M. It can easily be seen that β is smaller than unity if we consider an electron passing through at the moment $t_M = 0$.

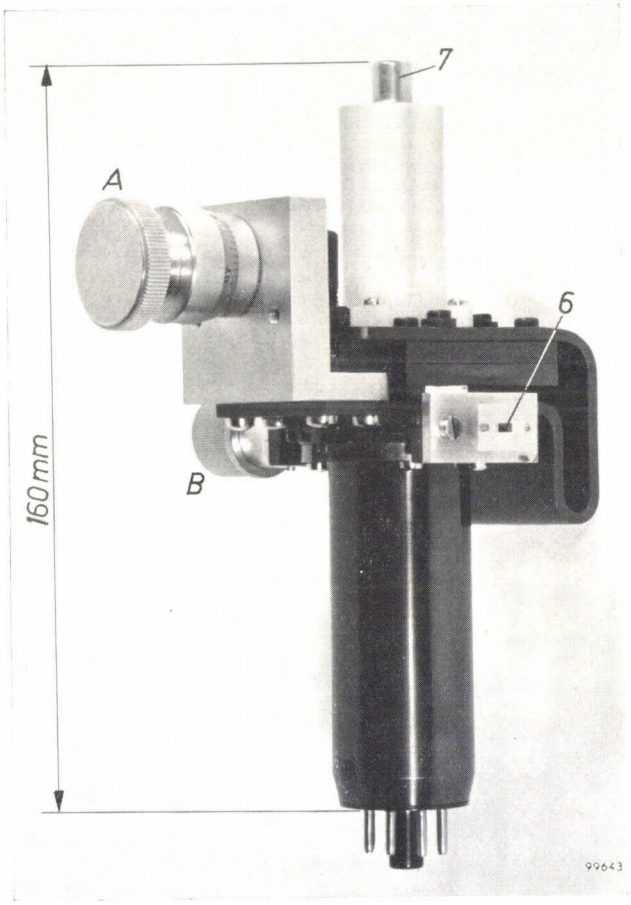


Fig. 6. Reflex klystron, type DX 151, for 4 mm waves, complete with tuning mechanism and other accessories. *6* output waveguide. *7* repeller connection. *A* tuning knob. *B* knob for adjusting shorting plunger. Principal electrical data:

Beam voltage V_0	2500 V
Beam current I_0	15-17 mA
Repeller voltage V_r	—150 to —500 V
Current density J_k at the cathode	2.4 A/cm²
Current density in the beam, in the interaction space	120 A/cm²
Power output P_o in the middle of the frequency band	100-200 mW
Mechanical tuning range	15 %
Electronic tuning range [4])	~100 Mc/s

[4]) The "electronic" tuning range is the range over which the frequency changes when V_r is varied between the two values at which the power has dropped to half its maximum value (the 3 dB points).

[5]) D. R. Hamilton, J. K. Knipp and J. B. H. Kuper, Klystrons and microwave triodes, Radiation Laboratory Series No. 7, McGraw-Hill, New York 1948, p. 311.

The bunching parameter X is equal to π times the product of the modulation depth $\tilde{V}/V_0$ and the electron transit time N. Where X is small, XI_0 (I_0 being the beam current) is exactly equal to the alternating current produced in the beam as a result of the velocity-modulation process [6]). At greater values of X saturation occurs, as a result of which the alternating current component becomes smaller than XI_0. The value which X (and hence the alternating voltage $\tilde{V}$) assumes in the oscillating tube depends not only on the tube itself but on the value of the externally applied load. For X we can write:

$$J_0(X) = \frac{G_R + G_B}{\pi k N \beta_1 \beta_2 G_0}, \quad \ldots \quad (2)$$

where

J_0 is a Bessel function of the first kind and zero order,

G_R is the resonance conductance of the resonant cavity,

G_B is the load conductance formed by the electron beam, and

G_0 is the D.C. conductance of the beam, i.e. I_0/V_0.

We shall first apply formulae (1) and (2) to the 4 mm klystron, type DX 151. The approximate relevant data for this type are: $k = 0.8$, $\beta_1 = 0.36$, $\beta_2 = 0.47$ and $N \approx 7$, hence $\pi k N \beta_1 \beta_2 \approx 3$. Taking into account that the resonant cavity rises to an average temperature of 175 °C above room temperature, we calculate for G_R the value

$$13 \times 10^{-6} \text{ ohm}^{-1}.$$

G_B is of the order of 0.1×10^{-6} ohm^{-1}, and is thus practically negligible compared with G_R, and where $I_0 = 15$ mA we have

$$G_0 = \frac{15 \text{ mA}}{2500 \text{ V}} = 6 \times 10^{-6} \text{ ohm}^{-1}.$$

From (2) we now find

$$J_0(X) = \frac{13 \times 10^{-6}}{3 \times 6 \times 10^{-6}} = 0.72.$$

The corresponding value of X substituted in (1) then gives $\eta = 0.5\%$. At a power consumption of 15 mA $\times$ 2500 V = 37.5 W, this corresponds to an output of about 200 mW, and with the best of the fabricated tubes this value is in fact achieved.

To produce a tube for still shorter wavelengths, the obvious thing to do is to scale down the dimensions in proportion to the wavelength. At this juncture we shall disregard the technical difficulties and increased dissipation per unit surface area which this involves. Scaling-down does not affect the quantities occurring in (1) and (2), with the exception of G_R, which increases proportionally with the root of the frequency. Taking the DX 151 as the tube to be scaled down, we find that for a given reduction factor the numerator of (2) becomes equal to the denominator, in which case $J_0(X) = 1$ and therefore $X = 0$. According to (1) the efficiency is then zero, that is to say, the tube no longer oscillates. The minimum wavelength at which this is the case is found to be 1.5 mm.

There are various reasons why this theoretical minimum wavelength cannot be achieved in practice. One possible reason, heat dissipation, has been mentioned above: the tube gets excessively hot if the surface area becomes too small. Another reason may be electrical breakdown, owing to insufficient clearance between the parts carrying the accelerating potential (in the DX 151 there is a potential of 3000 V between points only 0.17 mm apart!).

A third reason is that, at $\lambda = 1.5$ mm, the current density J_k at the cathode is greater than any known cathode is capable of delivering under conditions of space-charge limitation. The result of scaling-down is to decrease the surface area of the cathode in proportion to λ^2, although the current I_0 remains constant. Taking 4.3 mm as the average optimum wavelength of the DX 151, and $J_k = 2.4$ A/cm^2, which is the value of the current density for this tube, we find for J_k at $\lambda = 1.5$ mm:

$$\left(\frac{4.3}{1.5}\right)^2 \times 2.4 = 20 \text{ A/cm}^2.$$

A good L cathode cannot be expected to give more than 8 A/cm^2, so that as far as the cathode is concerned the minimum wavelength is

$$\sqrt{\frac{2.4}{8}} \times 4.3 = 2.4 \text{ mm}.$$

It might be asked whether a modified design would not make it possible to generate waves shorter than 2.4 mm without the current density on the cathode having to be greater than 8 A/cm^2. Such a design would in any case have to satisfy the oscillation condition. This condition, as far as the elementary theory applies, is that the right-hand side of (2) must be smaller than unity, i.e.:

$$G_R + G_B < \pi k N \beta_1 \beta_2 G_0. \quad \ldots \quad (3)$$

It is thus a matter of designing a tube which, at a wavelength as short as possible, will still satisfy (3) without J_k exceeding a specified value. A complete answer to this problem cannot be given since

[6]) See Philips tech. Rev. **13**, 210, 1951/52. The product $\pi a \xi$ found there is identical with the parameter X.

effects are involved that do not lend themselves to theoretical treatment. With the aid of experimental data, however, an estimate can be made of the minimum wavelength obtainable with a reflex klystron (but leaving out of account such questions as heat dissipation and electrical break-down). This estimate is arrived at by determining 1) the magnitude of the starting current, i.e. the minimum beam current needed according to (3) to excite oscillations, and 2) whether this current can be achieved.

The first step amounts to determining the starting current as a function of wavelength, the accelerating voltage V_0 and the diameter $2b$ of the holes in the resonant cavity. The remaining parameters, i.e. the other dimensions of the resonant cavity and the transit time N in the repeller space, must have the optimum values necessary to make the starting current as small as possible. Choosing the optimum value of N may be rather difficult. The oscillation condition, according to the elementary theory, implies that the starting current should constantly decrease as N increases and should show no minimum as a function of N. This theory, however, takes no account of various effects such as debunching of the electrons under the influence of space-charge forces, as a result of which the alternating current in the beam decreases, and differences in the transit time of electrons which describe different trajectories (transit-time spread). Above a certain value of N these effects cause the starting current to increase rapidly with N. Consequently, the starting current does in fact show a definite minimum as a function of N, but its position cannot be calculated. From experimental data of a number of differently-dimensioned tubes we found the best value to be $N \approx 8$. (It might be possible to obtian a more favourable value, however, by a more efficient design of the electrodes.)

The second question was whether the required starting current I_0 can be achieved, i.e. whether the electron beam with the current I_0, and at given values of J_k and V_0, can be made so narrow in cross-section as to allow it to pass through the resonator holes whose diameters are $2b$. It appears that in the range of parameter values with which we are concerned, the cross-section of the beam at the narrowest position is governed entirely by the initial thermal velocities of the electrons. This fact is demonstrated in the case of tube DX 151: calculation, disregarding the thermal velocities, shows 7 μ to be the smallest diameter of the beam, but the result when the thermal velocities are taken into account is 125 μ. The latter value agrees with the measurements.

Fig. 7 shows the results of the calculation for tubes in which $V_0 = 2000$ V and where the cathode temperature was 1450 °K (at which the L cathode delivers its maximum current density). The solid curves represent the beam current I_0, required to excite oscillations, as a function of the radius b of the holes, with the wavelength as parameter. The dashed lines represent the maximum beam currents that can pass through a hole having a radius b, plotted as functions of b for various values of J_k.

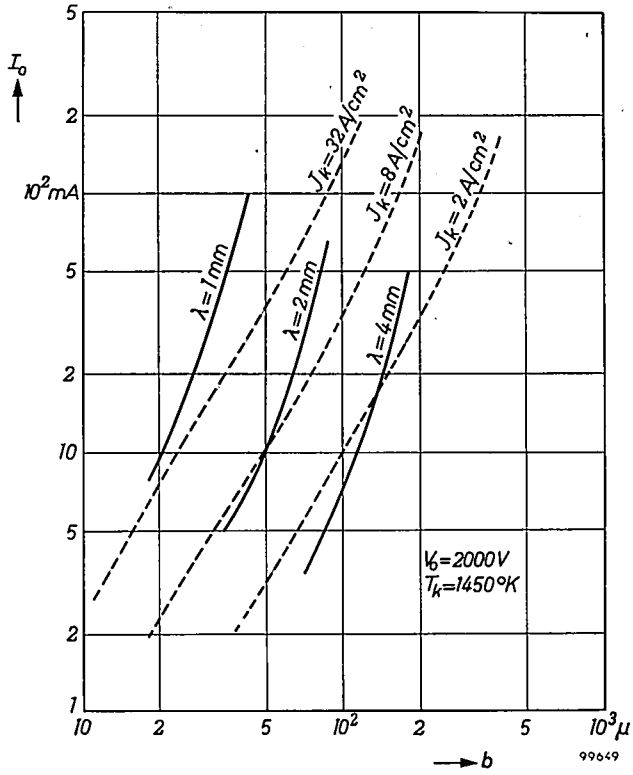


Fig. 7. Graph for calculating the minimum wavelength of reflex klystrons for $V_0 = 2000$ V, and a cathode temperature of 1450 °K. Solid curves: the calculated beam current I_0 required to excite oscillation, as a function of the radius b of the holes in the resonant cavity, for various wavelengths λ. Dashed lines: calculated maximum beam current I_0 capable of passing through a hole of radius b, as a function of b and with the current density J_k at the cathode as parameter. For $J_k = 8$ A/cm² the minimum wavelength is approximately 2 mm and the maximum value of $2b$ is 100 μ.

These calculations were confined to a Pierce electron gun. There is reason to assume, however, that other designs of gun will not yield appreciably better results. It was further assumed that the cone formed by the electron beam emerging from the gun has a semi-apex angle of $\frac{1}{3}$ radian, a value that will be difficult to exceed in tubes for wavelengths shorter than about 2 mm.

Oscillation at a given wavelength, and without J_k exceeding a specific value, can only occur at values of b for which the relevant solid curves in fig. 7 lie below the dashed curve corresponding to the per-

missible value of J_k. It can be seen from the figure that, at the maximum attainable emission density $J_k = 8$ A/cm² (and $V_0 = 2000$ V), the minimum wavelength is about 2 mm, for which the maximum permissible diameter $2b$ of the resonator holes is 100 μ. The minimum wavelength found from a similar graph for $V_0 = 4000$ V would be only slightly shorter, namely 1.6 mm. Electrical breakdown, however, would be much more difficult to avoid than at 2000 V.

A reflex klystron for 2.5 mm waves

Whether the theoretically estimated minimum wavelength can in fact be achieved depends on various circumstances already mentioned in passing, such as heat dissipation, electrical breakdown and constructional difficulties. It is not yet possible to say to what extent these problems will govern the minimum wavelength obtainable in practice. The results obtained with the tube discussed below have shown, however, that it is certainly possible to achieve a wavelength of 2.5 mm.

A tube for 2.5 mm waves cannot be produced by simply scaling-down the 4 mm type, since the heat dissipation would then be inadequate, at least in continuous operation. Nevertheless it proved possible — although perhaps somewhat at the expense of the electrical properties of the matching transformer — to design a suitable tube in which the heat can be adequately dissipated. The repeller presented some difficulties, since it was not possible to reduce its distance from the resonant cavity in view of the danger of breakdown. However, with suitably modified dimensions, the repeller continued to fulfil its purpose without having to be placed nearer to the resonator.

A cross-section of the tube is shown in *fig. 8*. The waveguides are now circular instead of rectangular, thus simplifying the construction of the mica windows and plunger. One of the windows (*18*) is so close to the annular space as to make the length of the waveguide to the plunger *19* particularly small. This reduces the losses in this section of waveguide, which adversely affect the properties of the matching transformer. Here too, the electron gun and the repeller are mounted in such a way as to permit after-alignment when the tube is in operation.

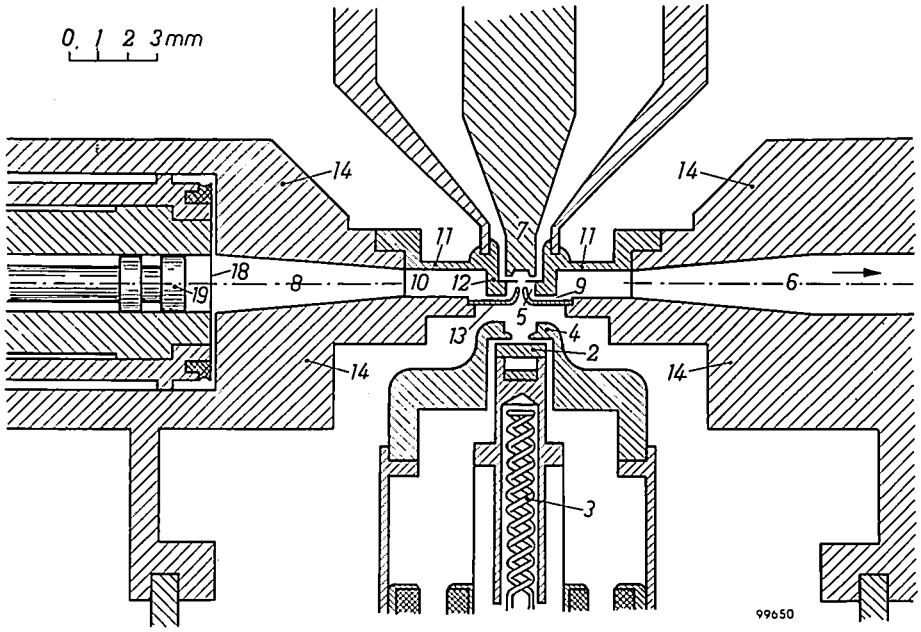


Fig. 8. Central portion of experimental reflex klystron for 2.5 mm waves, shown in axial section. Numbers *2* to *14* as in fig. 2. *18* mica window. *19* shorting plunger. Dimensions:

Diameter of opening in electrode *4* 0.6 mm	Diameter of holes in *12* and *13* . 0.15 mm
Diameter of resonant cavity *5* . . 1.0 mm	Diameter of waveguides *6* and *8* . 2.0 mm
Height of resonant cavity *5* . . . 0.4 mm	

The accelerating voltage is 2400 to 2500 V, the beam current 15 to 17 mA. The electron beam has a diameter of about 80 μ at its narrowest point, where the current density is slightly over 300 A/cm².

As with the 4 mm tube, the frequency can be varied mechanically over a range of 10 to 15% by varying the height of the resonant cavity.

Fig. 9 shows the output P_0 as a function of frequency and wavelength, measured on two experi-

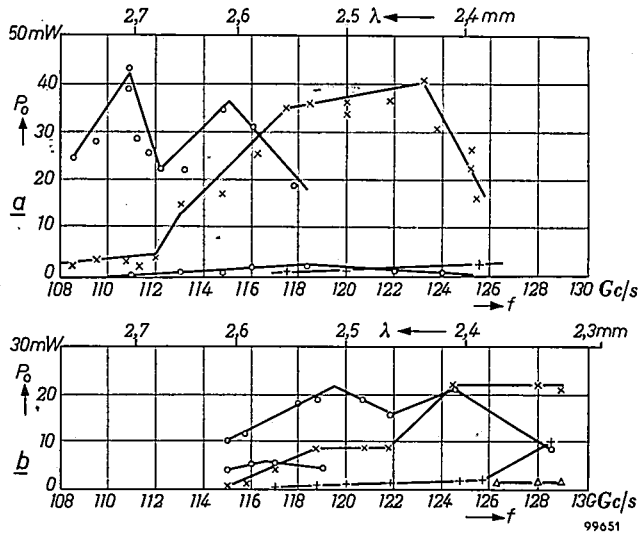


Fig. 9. Power output P_0 of two experimental reflex klystrons for 2.5 mm waves, as a function of frequency f and wavelength λ. The curves refer to various values of electron transit time N.

mental tubes. At $\lambda = 2.5$ mm an output of 40 mW was obtained; at $\lambda = 2.3$ mm — the shortest wavelength thus far generated by reflex klystrons — the measured output was 20 mW.

The curves in *fig. 10* represent the optimum repeller voltage V_r as a function of the mechanically adjusted frequency, for various values of the electron transit time N.

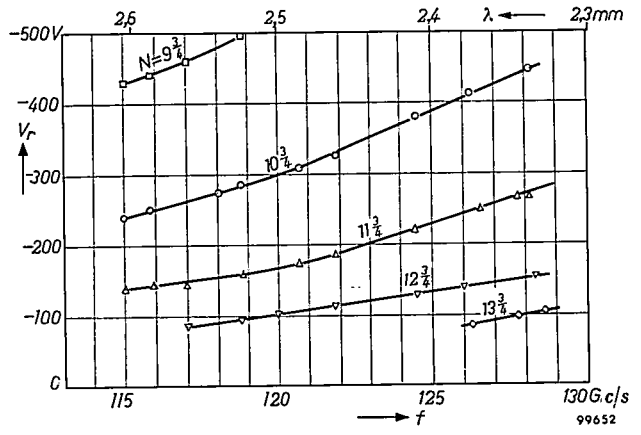


Fig. 10. The optimum repeller voltage V_r of one of the experimental 2.5 mm reflex klystrons, as a function of frequency f and wavelength λ, adjusted with the tuning mechanism. The curves relate to successive values of electron transit time N.

The output P_0 was measured by means of a bolometer employing a thermistor [7]). The thermistor was about 1 mm long, i.e. almost half a wavelength, and was mounted in a waveguide of 2 mm diameter. In spite of the relatively large dimensions of the thermistor, the error in the power measurement was calculated to be no more than a few per cent, and measurements have confirmed this. High-frequency losses occur outside the thermistor, however. The curves in fig. 9 were not corrected for these losses, and they therefore indicate the lower limit of the output.

[7]) In the mm wave region a bolometer using a thermistor has special advantages. For the temperature-dependent resistor of a bolometer there is generally a choice between a thin metal wire and a thermistor. A wire, long enough to dissipate sufficient power, will not be short enough compared with the wavelength in the case of mm waves. As a result, appreciable temperature differences will occur along the wire when heated by a high-frequency current. Owing to these temperature differences the sensitivity will differ from the D.C. sensitivity for which the instrument is calibrated. On the other hand the thermistor material (ceramic semiconductor) has a so much higher ratio of thermal conductivity to electrical conductivity that, in spite of unequal heating, no appreciable temperature differences occur. In this case, then, the D.C. calibration can still be relied upon in the mm wave range.

Summary. The reflex klystron for a wavelength of 4 mm, mention of which was made some years ago, is now being manufactured in small numbers (type DX 151). It operates on 2500 V, draws a current of 15-17 mA and delivers an output of 100-200 mW. The cathode is of the dispenser type (L cathode). After a short description of this tube, the question is considered of how far it is possible to go in generating even shorter waves with reflex klystrons. If the essential dimensions of an existing tube, e.g. type DX 151, are proportionately scaled down, various difficulties are soon encountered, such as excessive current density at the cathode, inadequate heat dissipation and increased risk of electrical breakdown. With 8 A/cm² as the maximum current density obtainable from an L cathode, and an accelerating voltage of 2000 V, the theoretical minimum wavelength is estimated at about 2 mm. Experimental tubes have been designed along these lines for a wavelength of 2.5 mm. They operate on 2400-2500 V and 15-17 mA; the output measured at 2.5 mm is about 40 mW and at 2.3 mm about 20 mW.

A SIMPLE CIRCUIT FOR A LIGHT SOURCE OF CONSTANT INTENSITY

by H. van SUCHTELEN. 535.241.42:621.3.078.3:621.383.4

In many measurements employing a light source, e.g. photometric measurements of luminous intensity, and absorption measurements with a colorimeter or spectrometer, it is important that the light source should be of very constant intensity. This article describes a simple circuit in which a cadmium-sulphide photoresistor [1]) is used to keep the emission of an incandescent lamp at constant intensity.

The luminous intensity of an incandescent lamp depends closely on the applied voltage. Curve *a* in *fig. 1* shows the percentage change in the luminous intensity of a given lamp in relation to voltage variations from a nominal 220 V. A 10% voltage variation changes the luminous intensity by 34%. For the purpose of measurements these voltage variations are usually eliminated by supplying the lamp from accumulators or via an A.C. voltage stabilizer. An accumulator can deliver a sufficiently constant voltage, but it requires a great deal of attention; moreover its bulk is a disadvantage. Simple stabilizers can be designed using a transformer with a saturated core [2]). The circuit shown in *fig. 2*, for example, provides a degree of stabilization. However, as can be seen from curve *b* in fig. 1, where the alternating supply was stabilized with this circuit, it is not adequate where an incandescent lamp of really constant intensity is required.

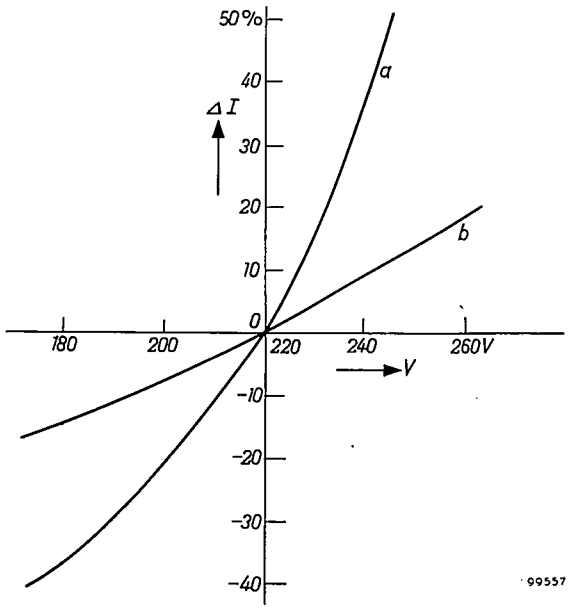


Fig. 1. Effect of mains voltage fluctuations on the luminous intensity of an incandescent lamp. The curves show the percentage change of luminous intensity in relation to the value at 220 V, *a* for a lamp without stabilizer, *b* for a lamp connected to a simple A.C. voltage stabilizer.

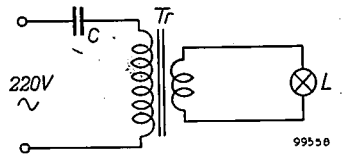


Fig. 2. Circuit of a simple mains-voltage stabilizer. *Tr* transformer with saturated core, *C* capacitor, *L* load.

Even if a better voltage stabilizer or a battery be used, however, there still remains the drawback that in the long run, the voltage being constant, the luminous intensity will decline due to the gradual evaporation of the filament and deposition on the bulb. Fundamentally, therefore, the luminous intensity *itself* should be stabilized rather than the supply voltage.

Although this can quite readily be done with an orthodox high-vacuum photocell and electron tubes, it has never found much practical application. The apparatus required is probably somewhat too complicated to be appreciated in the circles where it would be used in this case. Advances in the development of photoresistors and transistors have now made it possible, however, to design simple circuits that are eminently suited for stabilizing a small incandescent lamp of e.g. 30 watts. An experimental apparatus of this type has been built for use with a spectrometer in the Philips Research Laboratories, Eindhoven.

The circuit was designed for a 6V/5A car headlamp bulb. To start with, the circuit in fig. 2 was used, the mains voltage of 220 V having in any case to be stepped down. The voltage being thus roughly stabilized, several transistors were connected in series with the lamp. These transistors are connected in such a way that they decrease the current as the amount of light incident on a photoresistor increases. The principle of the circuit is illustrated in *fig. 3*.

The stabilizing transformer *Tr* has two secondary windings, S_1 and S_2, of which S_1 serves for the lamp supply. The current of approx. 5 A is regulated by

[1]) N. A. de Gier, W. van Gool and J. G. van Santen, Photoresistors made of compressed and sintered cadmium sulphide, Philips tech. Rev. **20**, 277-287, 1958/59 (No. 10).

[2]) H. A. W. Klinkhamer, A rectifier for small telephone exchanges, Philips tech. Rev. **6**, 39-45, 1941.

four type OC 16 transistors (T_2-T_5), connected in parallel in the common-emitter arrangement [3]). Since they pass the current in one direction only, they are bridged by a germanium diode D (type OA 31). The current through the lamp is therefore regulated only in one half of each cycle. An equalizing resistor of 1 Ω is connected in the emitter lead of each transistor in order to divide the current equally between them.

To explain how the regulation operates we shall first assume that the illumination is such that the bridge is balanced. If the illumination now increases, the value R_f of the photoresistor drops, causing the base of the amplifying transistor to go negative with respect to the emitter, and this transistor conducts. As a result the collector becomes less negative, and the same applies to the bases of the regulating transistors to which it is connected. From the transistor characteristics we may infer that this voltage increase cuts off the transistors as soon as the common collector voltage exceeds a certain value. Consequently, the current through the lamp is momentarily interrupted for a certain fraction of the active half of the cycle. The waveforms of the voltage across the incandescent lamp are represented in *fig. 5*.

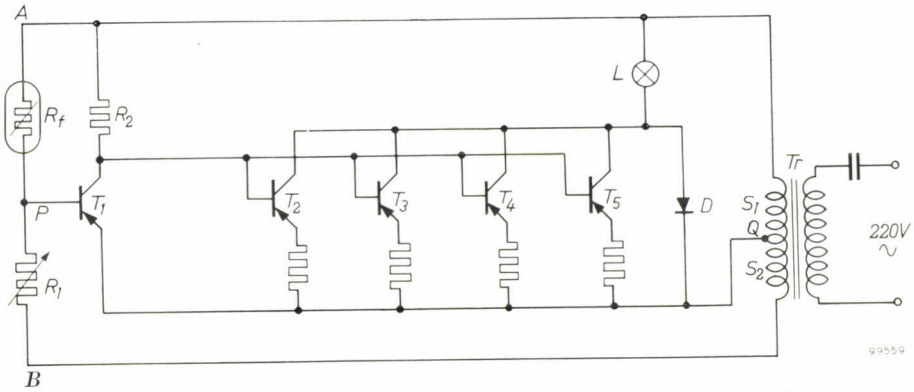


Fig. 3. Circuit for stabilizing luminous intensity. L incandescent lamp, R_f photoresistor, T_2-T_5 transistors regulating lamp current, T_1 transistor for amplifying control voltage.

Beside the lamp is mounted a cadmium-sulphide photoresistor (type ORP 30); it is enclosed in a can having a wedge-shaped aperture, which acts as a diaphragm for varying the illumination of the photoresistor (see *fig. 4*). This photoresistor (R_f in fig. 3) is included together with a variable resistor, R_1, in a bridge circuit, the other arm of which is formed by the secondary windings S_1 and S_2. Let the voltages in these windings be V_1 and V_2 and assume that they are in phase, then the bridge will be balanced (i.e. the voltage between P and Q will be zero) when $R_f : R_1 = V_1 : V_2$. This will be so only at one particular value of R_f, that is at a specific illumination of the photoresistor. A change in the illumination produces a voltage between P and Q. This voltage, the bridge output, is applied to the input of a transistor T_1, which acts as an amplifier and is again of type OC 16 in the common-emitter arrangement. The output voltage of T_1, developed across the resistor R_2, drives the four regulating transistors. The transistors are active only in that half of the A.C. cycle in which the collectors are negative; we need consider only this half cycle. (In the other half cycle the lamp is fed via the diode D.) In fig. 3, then, point A is negative with respect to point B.

In the inactive half cycle, the full current flows through the lamp via the diode. If the light source

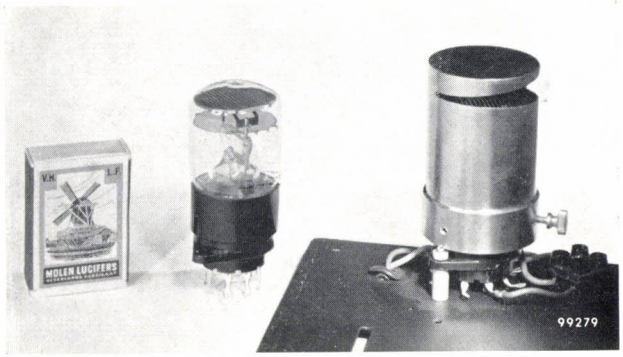


Fig. 4. Photoresistor, type ORP 30, mounted in a can with wedge-shaped aperture (diaphragm), seen from the light source. The light enters the aperture, where it is diffusely reflected so that every part of the horizontal photosensitive surface inside is equally illuminated. The illumination of the photoresistor can be varied by turning the can, which varies the effective width of the aperture. Left: the photoresistor shown separately.

[3]) J. P. Beijersbergen, M. Beun and J. te Winkel, The junction transistor as a network element at low frequencies, I, Philips tech. Rev. **19**, 15-27, 1957/58.

and photosensitive element were inertialess (as they would be if a discharge lamp and a high-vacuum photocell were used), the luminous intensity would accordingly not have the desired value during this

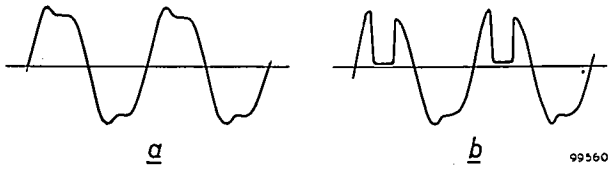


Fig. 5. Voltage waveform across the incandescent lamp, *a*) connected only to A.C. voltage stabilizer (which is responsible for the distortion of the sine wave), *b*) incorporated in the circuit described.

half cycle. The regulation would then deal with only half the effect of a voltage variation. In fact, however, both the lamp and the photoresistor possess inertia. The luminous intensity does not follow the momentary change of the mains voltage, but remains at a value which is the average over several cycles. The regulation in the active half cycle is thus effective for voltage variations in the other half too.

This control system does not lend itself well to a mathematical approach. The behaviour of the circuit can be very simply determined, however, by varying the magnitude of R_f in the absence of any feedback from the lamp to the photoresistor. For this purpose the photoresistor is replaced by a fixed resistance in series with a small variable resistance; the luminous intensity of the source is then measured as a function of R_f, at various voltages. What is in fact measured is not the luminous intensity I but the illumination E in a plane V perpendicular to the beam of light at the position of the photoresistor diaphragm (see *fig. 6*). Since luminous intensity and illumination are proportional to one another, we shall use E instead of I in the following considerations. The measured curves of $E = \varphi(R_f)$ are shown in *fig. 7*. It can be seen that a relatively slight decrease of R_f is indeed accompanied by a fairly considerable drop in illumination. The effect of the mains voltage in the absence of feedback is also apparent (the stabilizing action of the mains transformer is implicit in the curves).

We can now represent the feedback graphically by plotting the line which indicates the value of R_f as a function of E. The characteristic of the photoresistor [4]) is such that R_f is inversely proportional to the illumination on the photosensitive surface. The value of this illumination is less than the above-mentioned E in the plane V, but the two values stand in a fixed relation to one another at any fixed position of the diaphragm. Hence R_f is also inversely proportional to E, and the relevant curve is part of a rectangular hyperbola $ER_f =$ constant. At a given setting of the diaphragm we

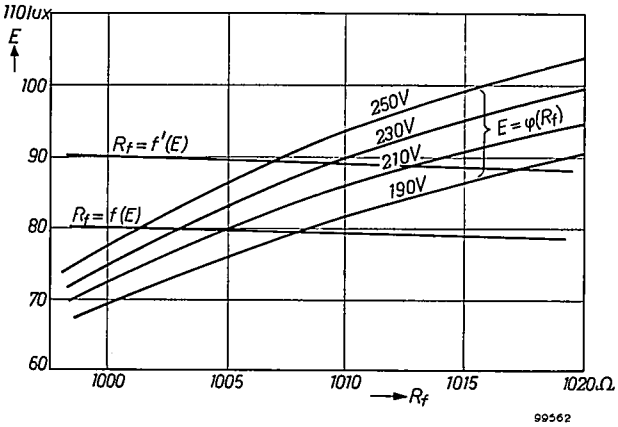


Fig. 7. Circuit characteristics, where E on the ordinate is the illumination at a fixed position just in front of the photoresistor diaphragm.

The curves $E = \varphi(R_f)$ are the characteristics of the circuit in the absence of feedback from lamp to photoresistor. The lines $R_f = f(E)$ and $R_f = f'(E)$ give the relation between E and R_f for the photoresistor, at two different positions of the diaphragm. At a given mains voltage the operating point of the circuit is the point of intersection between the relevant φ curve and f-line.

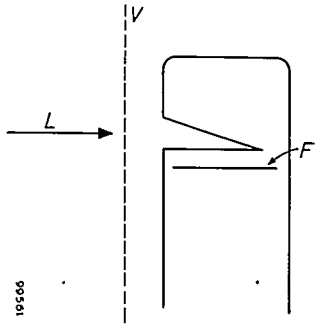


Fig. 6. Schematic cross-section of photoresistor with diaphragm. F photosensitive surface, L direction from which light enters diaphragm, V plane in which the illumination level is measured (E in fig. 7).

obtain the line $R_f = f(E)$ represented in fig. 7. The operating point of the circuit at a particular mains voltage is now the point where this line $R_f = f(E)$ intersects the measured curve $E = \varphi(R_f)$. If the mains voltage changes, this point shifts along the line $R_f = f(E)$. Between 190 V and 250 V we see that E varies from about 79.5 to 80 lux, i.e. by about 0.5%.

By changing the position of the diaphragm a different luminous-intensity setting is obtained. If we require a higher luminous intensity, for example, we turn the can (fig. 4) so that less light enters the

[4]) See fig. 7 in article [1]).

diaphragm. Whilst E in plane V remains the same (fig. 6), the illumination on the photosensitive surface is now reduced. The resistance therefore increases, and the bridge is unbalanced. Point P in fig. 3 is now more positive, less current flows through T_1 and more current is passed by the regulating transistors. As a result, E increases, so that R_f again decreases until a new balance is reached. In this state of balance R_f will remain higher than before the change, inasmuch as more current must now flow through the lamp.

This can also be seen from fig. 7. If we consider only the line $R_f = f(E)$, distinct from the circuit, we see that it changes position when the diaphragm is reduced. Since R_f is larger for the same value of E, the line shifts to the right and is higher in the diagram than the original; see the curve $R_f = f'(E)$. The φ curves remain unchanged, so that the operating point at a given mains voltage now lies at a higher value of E. The value of R_f is also higher.

However, E now varies more than it did before with a given mains voltage fluctuation. Along the f-line the percentage change of E is equal to that of R_f (evidently, since differentiation of $ER_f = \text{const.}$ yields $dE/E = -dR_f/R_f$). The same applies to the f'-line. Upon a given change of R_f, then, $\Delta E/E$ is the same for both curves. An increase of E therefore means a proportional increase of ΔE, that is to say, a proportionally steeper slope of the line. Since the φ curves are roughly parallel with one another, it may be said that the variations referred to are proportional to E.

The circuit should be so adjusted as to allow the luminous intensity to be regulated for mains fluctuations in both directions. At 220 V, therefore, a part of each cycle should be suppressed, e.g. as shown in fig. 5*b*. For this purpose P in fig. 3 should be given a certain negative potential, which is done by varying R_1. Thus, in changing over to a different E setting, it is necessary to adjust both the diaphragm and R_1.

We shall now consider to what extent temperature variations affect the stability of the system. First, the effect on the photoresistor. The resistance of the CdS photoresistor increases by about 0.2% per °C temperature increase at 20 °C [5]). This means that the f-line in fig. 7 is shifted 2 Ω to the right (0.2% of 1000 Ω). The φ curve corresponding to the nominal mains voltage remains unchanged; the operating point therefore moves to a higher position on this curve. If the φ curve were vertical, E would also increase by 0.2% (see above for f-line). Although the increase of E is somewhat smaller, the difference is negligible because the slope of the φ curves is much steeper than that of the f-lines. The increase in illumination when the temperature of the photoresistor rises by 1 °C can accordingly be put at 0.2%.

Temperature variations also affect the transistors; the most sensitive in this circuit is the amplifying transistor T_1. In an earlier article on transistors [6]) it was shown that a temperature increase of 1 °C has the same effect as a voltage drop of 2.3 mV at the base (point P in our case). As a result the collector current rises, the collector goes more positive and the current through the lamp drops. This temperature effect is thus the opposite of that on the photoresistor.

The reduction of E can be calculated most simply by using the same reasoning as applied to the temperature effect on the photoresistor. The voltage drop at P, caused by the increased temperature of the transistor, can then be regarded as due to a drop in the value of the photoresistor. As in the preceding case, the φ curve then remains unchanged, but the f-line now shifts to the left. Here too, and for the same reason as above, the percentage change of E is roughly equal to that of R_f. We therefore have to find the decrease in R_f that causes a voltage drop of 2.3 mV at P. This follows directly from the formula for the output voltage V_o of a Wheatstone bridge with identical resistances, one of which is varied:

$$V_o = \tfrac{1}{4} V \frac{\Delta R}{R},$$

where V is the bridge supply voltage (15 V). For $V_o = 2.3\text{m V}$ we find a reduction $\Delta R/R$ of 0.06%, so that the decrease in E per °C increase in the transistor temperature may also be put at 0.06%.

It might be thought from the above that the two temperature effects would oppose each other, and that therefore the deviation should to some extent be compensated. In fact, the photoresistor and transistors do not undergo the same temperature variations, since they are in separate compartments in the actual apparatus. The most unfavourable case, i.e. oppositely changing temperatures, is very unlikely, however. Under the influence of the ambient temperature they are much more likely to change in the same direction. We are therefore on the safe side if we disregard the smaller of the two effects and reckon on a 0.2% increase per °C.

[5]) See fig. 8 in article [1]).

[6]) J. P. Beijersbergen, M. Beun and J. te Winkel, The junction transistor as a network element at low frequencies, III, Philips tech. Rev. **20**, 122-134, 1958/59 (No. 5).

Assuming temperature variations of 5 °C, this means light variations of 1%.

Because of these temperature effects the circuit is not constant enough to be used as a light standard, particularly since no account has been taken of phenomena associated with the aging of the photoresistor. However, for use with a colorimeter or spectrometer, where stability is required only during successive measurements on test and comparison samples, the influence of temperature is no objection.

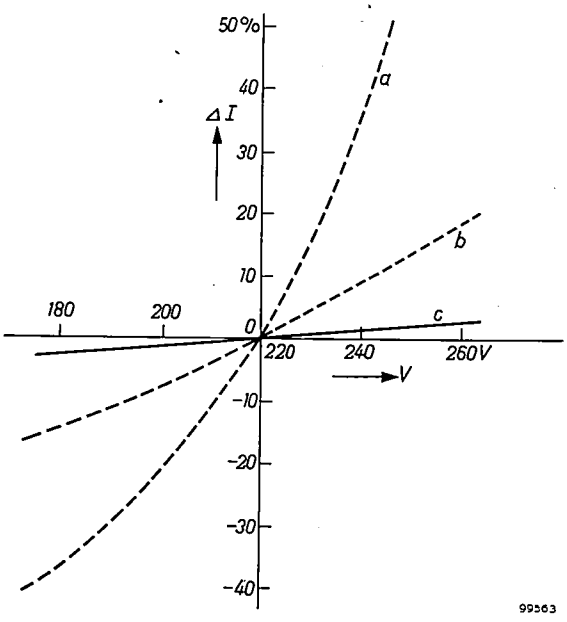


Fig. 8. Variation of the luminous intensity of an incandescent lamp with varying mains voltage. The curves show the percentage change of luminous intensity in relation to the value at 220 V, *a*) for lamp without stabilizer, *b*) for lamp fed by a simple voltage stabilizer as in fig. 2, *c*) for lamp incorporated in the circuit described. The curves *a* and *b* are those in fig. 1.

The photoresistor is mounted together with the lamp in one small compartment. Although this is ventilated, allowance must still be made for a warming-up period in which the luminous intensity increases. According to calculations the increase might be a few percent, and observation has also confirmed this. The luminous intensity has settled down to a constant value after about one hour, although the apparatus is stable enough for use after the first 20 minutes or so. A similar warming-up period is also prescribed for the spectrometer itself, for which the apparatus was primarily designed.

The apparatus is used at a luminous intensity twice as high as that at which the measurements for fig. 7 were done. On our reasoning this means that the variation of intensity with voltage should likewise be twice as high. The change, measured over a voltage excursion from 190 to 250 V, was in fact 1% (*fig. 8*).

Summary. A light source of constant intensity is commonly obtained by supplying an incandescent lamp from accumulators or a stabilized power supply. It is more logical, however, to stabilize the luminous intensity itself. A circuit designed for this purpose is described. The light source is a 6 V, 5 A incandescent lamp. The mains voltage of 220 V is first stepped down and stabilized by a transformer with a saturated core. A photoresistor in conjunction with five transistors regulates the luminous intensity. The latter remains constant to within 1% for mains voltage fluctuations of 20% in either direction. The effect of temperature variations is discussed; the intensity varies less than 0.2% per °C change in temperature.

FLYING-SPOT SCANNERS FOR COLOUR TELEVISION

by H. van GINKEL.

621.385.832:621.397.611.2

Flying-spot scanners for colour television are of interest chiefly because, in principle, their signals are completely free from errors of superposition or "register". The article below describes two types of flying-spot scanner that have rendered valuable service in development work on colour television at Philips. One is suitable for colour slides, the other for opaque objects (colour prints, drawings, small objects etc.). Both types probably have a useful part to play in colour television broadcasts.

Methods of generating primary colour signals

In colour television three camera tubes are used at the transmitting end, one for each of the three primary colours (red, green and blue) into which the incident light is split by special devices. The principles of such systems have been dealt with at some length in an earlier article in this journal [1]) which will henceforth be referred to as I. The three primary colour signals thus obtained are transmitted, as described in I, to the receiver. Here they are combined by a process of additive mixing to produce a multi-coloured picture, either on the fluorescent screen of a special direct-viewing tube, or on a projection screen on to which the three primary-colour images are projected [2]).

In the development of the various circuits of a colour television system it is obviously necessary to have a signal source for the three primary-colour signals. These can be obtained from an electronic signal generator, a colour-television camera or a flying-spot scanner. The advantages and drawbacks of these three methods are mentioned below.

a) *Electronic signal generator.* The signals represented in *fig. 1a* can be generated fairly simply. If they are used as "red", "green" and "blue" signals, respectively, a combination of them produces a television image consisting of variously coloured vertical strips (fig. 1*b*). This method, although it has the virtue of simplicity, is limited by its very primitive picture to only a few applications, such as testing certain properties of transmission systems.

b) *Studio camera(s).* The advantage of obtaining the primary-colour signals from studio cameras is that the situation is technically identical with an actual transmission and that the pictures can be given movement and detail. Apart from being more complicated, however, this method has the drawback that the signals delivered by a camera are subject, for reasons we shall explain, to slight errors. In experimental work this is a fundamental difficulty, for if the picture is unsatisfactory it is not usually known whether the cause lies in the part under investigation (transmission system, receiver, projector) or in the picture signals themselves.

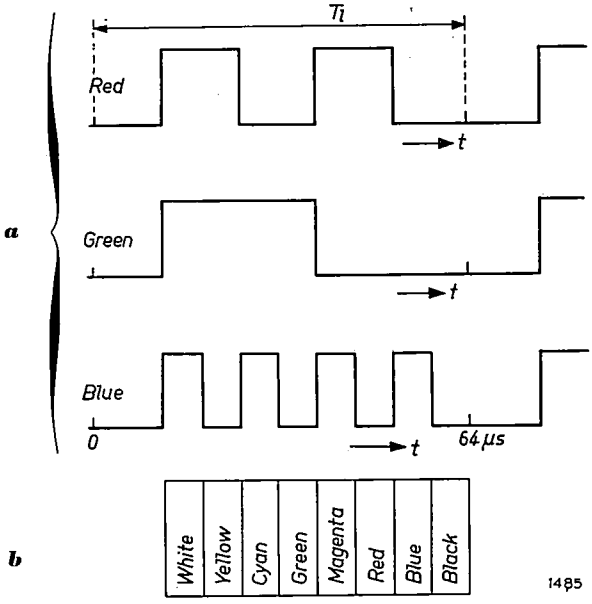


Fig. 1. *a*) Waveforms of primary-colour signals "red", "green" and "blue", produced by signal generators. T_l is the period of one complete line (scan plus flyback).
b) Combination of the three signals (*a*) produces a television picture consisting of vertical strips of the colours indicated.

Fig. 2 illustrates the path of rays in a camera containing three vidicons. As explained in article I (fig. 7) the light entering the camera is separated by dichroic mirrors [3]) into three spectral components, red, green and blue, each of which is picked up by a separate vidicon.

[1]) F. W. de Vrijer, Fundamentals of colour television, Philips tech. Rev. **19**, 86-97, 1957/58.
[2]) T. Poorter and F. W. de Vrijer, The projection of colour-television pictures, Philips tech. Rev. **19**, 338-355, 1957/58.
[3]) P. M. van Alphen, Applications of the interference of light in thin films, Philips tech. Rev. **19**, 59-67, 1957/58.

Because of light losses in the system, use is made only of highly sensitive camera tubes, such as image orthicons [4]) and vidicons [5]). The three primary-colour signals, one from each of the camera tubes, must be perfectly synchronized, i.e. the scanning

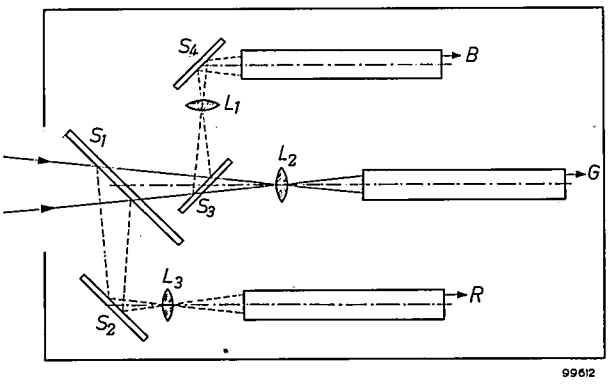


Fig. 2. Arrangement of the principal optical components and the vidicons in a colour-television camera, seen from above. R, G, B vidicons for red, green and blue, respectively. S_1, S_3 dichroic mirrors. S_2, S_4 ordinary mirrors. Light from the scene enters the system from the left. The red component is reflected by mirrors S_1 and S_2 and reaches the vidicon R; the blue component passes through S_1 and reaches the vidicon B via S_3 and S_4; the green component passes through both mirrors S_1 and S_3 and reaches vidicon G. The lenses L_1, L_2 and L_3 image the scene on the photoconductive layer of the vidicons. The images on the three vidicons and the individual scannings must be identical, otherwise some of the colour areas in the picture will show tinted edges.

beams in the three vidicons must pass corresponding image points at every instant. To meet this requirement it is essential to use identical objective lens systems and, by means of mirrors, to arrange the tubes in such a way that optically the three photoconductive layers apparently coincide. Further, the deflections — both horizontal and vertical — must have exactly the same amplitude, and even the unavoidable deviations from linearity (proportionality with time) should be equal. The latter is never entirely possible because there is always some spread in the properties of the components used; here, then, problems arise of the kind discussed in the article under [2]). The result of the failure to satisfy completely the conditions mentioned is that the three primary-colour images are out of register, i.e. they do not exactly coincide. Even if the reproduction system were faultless, the errors of register would still give rise to coloured edges between fields of different hue. The total picture then resembles a colour print on which the blocks were out of register during printing.

c) Flying-spot scanner. A flying-spot scanner is used for generating a television signal from a flat object, such as photographic transparencies, prints, drawings, etc. An article on a flying-spot scanner for black-and-white slides appeared in this journal some years ago [6]). Its operation is recalled in *fig. 3*.

Fig. 4 shows how the flying-spot scanner can be extended to make it capable of generating three colour signals; the figure also incorporates a block diagram of the electronic equipment required. The dichroic mirrors are here in a cruciform arrangement (cf. [3]), fig. 12). The red, the green and the blue components are each picked up by a corresponding photomultiplier tube. Across the load resistor of each tube there thus appears a picture signal voltage, which is amplified and, as explained below, corrected for the afterglow of the phosphor and the non-linear relation between the luminous flux and the control voltage of the picture tube.

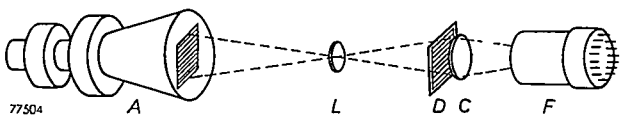


Fig. 3. Principle of a flying-spot scanner for black-and-white slides. The luminous spot on the cathode-ray tube A describes a line raster of constant brightness which is transmitted by the objective lens L on to the slide D. The condenser lens C collects the light transmitted by D and directs it on to the photocathode of the multiplier tube F. This luminous flux is proportional to the transmission of the transparency at each point scanned.

Since, in the system sketched, there is only one light spot on the cathode ray tube, and therefore only one light spot on the transparency, the three-primary-colour beams are perfectly synchronized. The same applies, then, to the three primary colour signals, and this is an important, fundamental advantage of the flying-spot scanner over the studio camera. Moreover, the flying-spot scanner is simpler in construction than a camera.

The principle of spot scanning need not be restricted in application to transparencies; it can equally be used for opaque matter such as prints and drawings.

In development work on colour television, wide use has been made of flying-spot scanners. They are indeed still employed at Philips for various purposes, as for example in factory inspection of

[4]) See e.g. A. Rose, P. K. Weimer and H. B. Law, The image orthicon — a sensitive television pickup tube, Proc. I.R.E. **34**, 424-432, 1946.

[5]) P. K. Weimer, S. V. Forgue and R. R. Goodrich, The Vidicon photoconductive camera tube, R.C.A. Rev. **12**, 306-313, 1951. See also L. Heijne, P. Schagen and H. Bruining, An experimental photoconductive camera tube for television, Philips tech. Rev. **16**, 23-25, 1954/55, and L. Heijne, The lead oxide vidicon, Acta Electron. **2**, 124-131, 1957/58.

[6]) F. H. J. van der Poel and J. J. P. Valeton, The flying-spot scanner, Philips tech. Rev. **15**, 221-232, 1953/54.

direct-viewing tubes for colour television. In the following section we shall discuss a scanner for transparencies (colour slides), evolved in the Philips Research Laboratories [7]) and further developed by the ELA Division of Philips; in the last section a laboratory model of a scanner for opaque matter will be described.

Flying-spot scanner for colour slides

The scanning tube

The flying-spot scanner for black-and-white slides, described in article [6]), uses a cathode-ray scanning tube of the type that was discussed in the same issue [8]). The tube used for colour slides differs only in one respect — the spectral emission of the phosphor. In the former case it was simply a matter of generating radiation which would be transmitted with adequate contrast by the slide and to which the photomultiplier tubes were sensitive; apart from this the spectrum of the radiation is immaterial, unless of course colour slides are to be reproduced in a monochrome transmission. In colour television, however, the emitted light is required to possess all wavelengths of the visible spectrum; each point of the slide then transmits a certain proportion of some colours and thus absorbs a complementary proportion of the other colours.

Apart from the spectral distribution of the radiation, the luminous efficiency η and the afterglow time τ of the phosphor have to meet specific requirements. (η is the ratio of the radiant power emitted to the electrical power consumed, and τ is the time in which the emitted light decays to a fraction 1/e of its initial value after the sudden removal of the excitation.) The ratio $\sqrt{\eta}/\tau$ can be used as a figure of merit for the phosphor (see [6])) in the sense that, if its value is raised, the system comprising phosphor and photomultiplier will give a better signal-to-noise ratio. Among the phosphors used for colour television the best is found to be ZnO containing an excess of Zn. As can be seen from the spectral characteristic in *fig. 5*, this phosphor emits green light with a maximum at a wavelength of 505 mμ. Elsewhere the kind of ZnO phosphor used in colour TV is one where $\eta = 8\%$ and $\tau = 2.5$ μs. For our scanning tubes we prefer the kind having a luminous efficiency of 2.5% and an afterglow time of 1 μs. This is a somewhat longer afterglow than that of the gehlenite phosphor used for monochrome television. The consequences will be discussed presently.

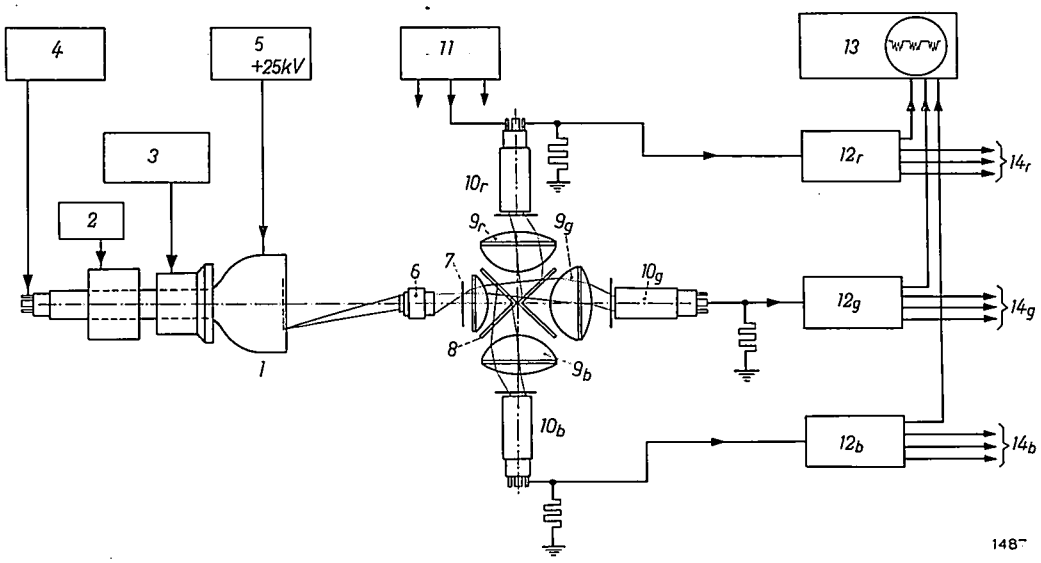


Fig. 4. Block diagram and optical system of a flying-spot scanner for colour slides. *1* scanning tube, *2* focusing coil, *3* generators for horizontal and vertical deflection, *4* stabilization of beam current and beam suppression during the flyback, and *5* anode-voltage generator (25 kV). *6* objective lens. *7* colour slide. *8* dichroic mirrors. *9* condenser lenses (the suffixes *r*, *g* and *b* relate to the "red", "green" and "blue" channels). *10* photomultiplier tubes (with filter in front of photocathode) and associated power pack *11*. *12* video channels with afterglow compensation, gamma correction and output amplifier. *13* test circuit with oscilloscope. *14* cables to black-and-white monitor, colour monitor and various signal consumer circuits.

[7]) The development work on this project was initiated by A. Wieberdink.

[8]) A. Bril, J. de Gier and H. A. Klasens, A cathode-ray tube for flying-spot scanning, Philips tech. Rev. **15**, 233-237, 1953/54.

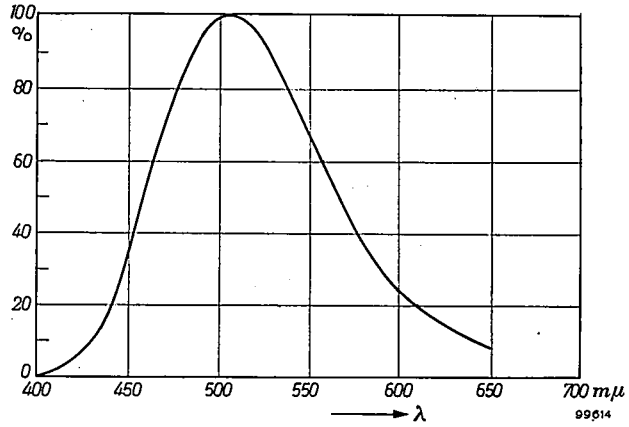


Fig. 5. Spectral characteristic of the light emitted by the ZnO-Zn phosphor employed in the scanner tube, measured in the Philips laboratories. Efficiency 2.5%. Afterglow time 1 μs.

Fig. 6 shows a scale drawing of the tube. The raster may be a maximum of 120 mm diagonally. The window is flat and must be entirely free of smudges and scratches, which would be imaged fairly sharply on the transparency. The electron gun is a triode type with an anode potential of 25 kV; as a rule the beam current is 150 μA. Magnetic focusing is employed.

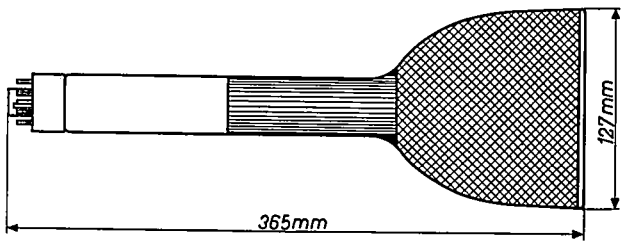


Fig. 6. Scale drawing of scanning tube.

The optical system

The path of the rays from tube window to photomultiplier cathode was illustrated in fig. 4. The major components of this optical system are the cruciform mirrors, the condenser lenses and the objective lens. To minimize light losses at the intersection of the mirrors (8 in fig. 4) the front mirrors, i.e. those facing the objective lens, are made thicker than the rear ones [9]).

The condenser lens system serves a dual purpose. Firstly, the condenser lens at the transparency side ensures that the beams emanating from points outside the middle of the raster are bent towards the axis of the system. This is done in such a way that the axes of these beams all reach the dichroic mirrors at virtually the same angle — in this case 45° — irrespective of their point of origin on the raster. This is necessary because the reflection coefficient of the mirrors is a function of wavelength, and hence the colour of the reflected light depends on the angle of incidence. Secondly, the condenser system serves to image the diaphragm of the objective lens on the photocathode (we shall return to this later). This is done by the lens at the transparency side in conjunction with the lens adjacent to the photocathode.

The objective lens is an anastigmat of aperture *f*. 2. When the flat raster on the tube, measuring 6×8 cm, was imaged with a reduction of 2.85 times on a colour slide measuring 21×28 mm, the definition at the edges proved to be perceptibly poorer than in the middle of the picture, even with smaller apertures. Improvement was effected by adding an auxiliary lens of +2.5 diopters.

Conversion of light into electric signal

As mentioned above, the light is converted into electric signals by means of photomultiplier tubes. The signal voltage delivered by these tubes (in our case 35 mV peak-to-peak) is high enough to discount the noise added by the amplifier.

One of the types of photomultiplier used in the flying-spot scanners is the Philips type 50 AVP photomultiplier tube [10]); a cut-away view of the tube appears in *fig. 7*. Each stage of this tube amplifies the current by a factor δ, the secondary emission coefficient. There are in all eleven stages, giving a total amplification of δ^{11}, which may be as high as 1 million times.

Because δ is raised to such a high power, the total amplification depends strongly on δ, and in its turn δ depends on the material of the dynodes and on the potential between successive dynodes. Fluctuations in these potentials therefore cause the amplification to vary. In monochrome television this gives rise only to variations of picture brightness. In colour television, however, there will in addition be colour variations if the amplification of the three photomultipliers varies unequally which, owing to unavoidable differences between individual tubes, is always the case. Variations in colour are much more troublesome than variations in brightness. The voltage of the source from which the dynodes receive their supply voltage must therefore be kept highly constant. Voltage-stabilizer tubes are used for this purpose.

[9]) See page 67 of the article under [3]). The idea is due to H. de Lang.

[10]) Also frequently used in scintillation counters; see Philips tech. Rev. **20**, 209-219, 1958/59 (No. 8).

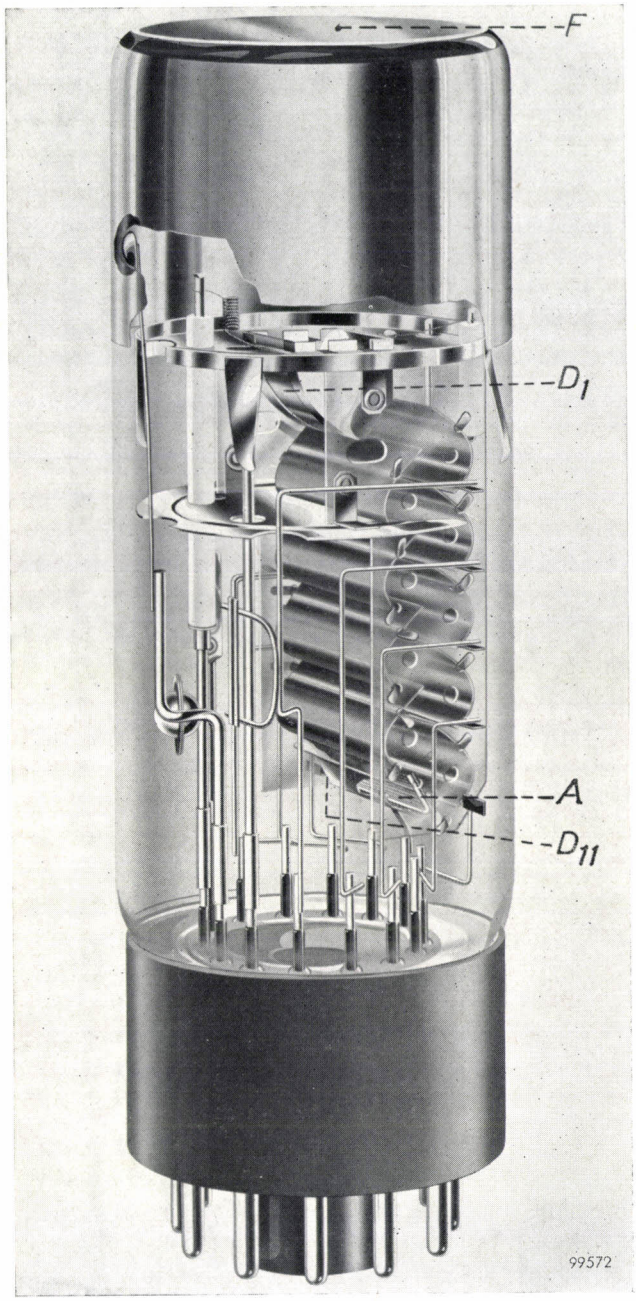


Fig. 7. Philips photomultiplier tube type 50 AVP. F photocathode. D_1 first dynode. D_{11} last dynode. A anode.

In general, the sensitivity of a multiplier tube is not uniform over the whole photocathode area. Furthermore, equal numbers of electrons issuing per second from various points of the photocathode will not always give rise to equal anode currents, owing to the fact that the first dynode is not an ideal collector. For these reasons, if the optical system were designed so as to image the slide itself on the photocathode, the resultant picture signal would show some slight modulation causing troublesome local colour differences to appear in the received picture. For this reason the diaphragm aperture, and not the slide, is imaged on the photocathode as mentioned earlier. The light originating from the luminous spot, and which has passed through the slide, is thus entirely "smeared out" over the part of the photocathode actually used.

The spectral sensitivity of the three channels

As remarked earlier, the scanning tube is required to emit radiation in the entire visible spectrum. This radiation, in so far as it is transmitted by the slide, is separated by the dichroic mirrors into red, green and blue components, which are picked up by the corresponding photomultiplier tubes. It was pointed out in article I that each of the channels thus formed must possess a very specific spectral sensitivity. The curves $\bar{R}(\lambda)$, $\bar{G}(\lambda)$ and $\bar{B}(\lambda)$ in *fig. 8* can be interpreted as representing the anode currents of the respective photomultiplier tubes for light entering the optical system of continuously changing wavelength and of constant intensity. The parts of the curves below the λ-axis are usually disregarded for colour television (see I, page 92).

It is found that the use of dichroic mirrors as selective elements does not directly produce these curves, and for this reason a correction filter is introduced into the path of the rays, usually immediately in front of each photocathode. The spectral transmission $D_f(\lambda)$ of such a filter, e.g. $D_{fg}(\lambda)$ of the green filter, is calculated from the desired sensitivity $\bar{G}(\lambda)$ and from the spectral energy distribution $F(\lambda)$ of the phosphor, the spectral

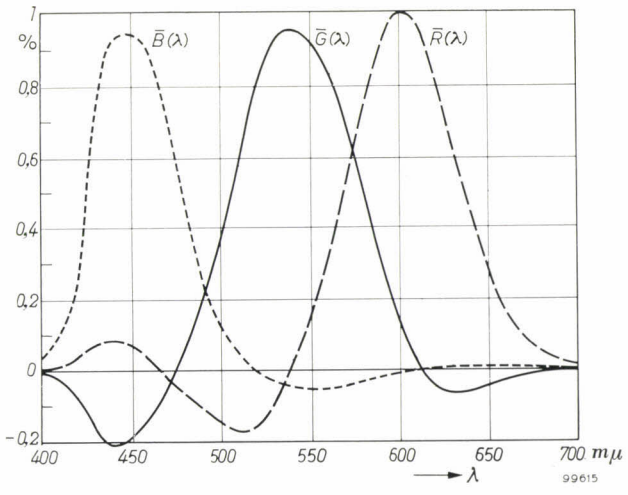


Fig. 8. Required spectral sensitivity of the three colour channels of a colour television system using primary colours whose colour points have the following coordinates:

red . . . $x = 0.67$, $y = 0.33$;
green . . . $x = 0.21$, $y = 0.71$;
blue . . . $x = 0.14$, $y = 0.08$.

Colourless objects are reproduced in standard white C, having the coordinates $x = 0.310$, $y = 0.316$.

transmissions $D_r(\lambda)$ and $D_b(\lambda)$ of the red and blue reflecting mirrors, and the spectral sensitivity $M_g(\lambda)$ of the photomultiplier tube in the "green" channel:

$$D_{fg} = \frac{\overline{G}(\lambda)}{F D_r D_b M_g}.$$

All quantities in this expression are functions of λ. The desired transmissions D_{fb} and D_{fr} in the "blue" and "red" channels, respectively, are similarly found from:

$$D_{fb} = \frac{\overline{B}(\lambda)}{F D_r R_b M_b}$$

and

$$D_{fr} = \frac{\overline{R}(\lambda)}{F D_b R_r M_r}.$$

Here R_b is the spectral reflection of the blue-reflecting mirror and M_b the spectral sensitivity of the photomultiplier tube in the "blue" channel, whilst R_r and M_r are the corresponding quantities for the red-reflecting mirror and the photomultiplier tube in the "red" channel.

Fig. 9 shows the spectral characteristics of the cruciform arrangement of dichroic mirrors relating to each of the channels, i.e. D_bR_r for the "red" channel, D_rD_b for the "green" channel and D_rR_b for the "blue" channel. In *fig. 10*, curve *1* represents the spectral sensitivity $F(\lambda)$ of the type 50AVP photomultiplier tube. It can be seen from fig. 8 that the maximum of $\overline{B}(\lambda)$ lies in a region where the type 50 AVP multiplier still has 95% of its peak sensitivity. This tube is therefore used in the "blue" channel.

The maximum of $\overline{R}(\lambda)$ is found at $\lambda = 600$ mμ, and here the spectral sensitivity of the 50 AVP tube has dropped appreciably. From the spectral energy distribution of the phosphor used (see fig. 5) we find moreover that, compared with green, very little red light may be expected. Since the signal-to-noise ratio of the photocurrent is proportional to the root of the number of primary electrons per picture element, the "red" photocurrent will therefore have a poorer signal-to-noise ratio than the "green".

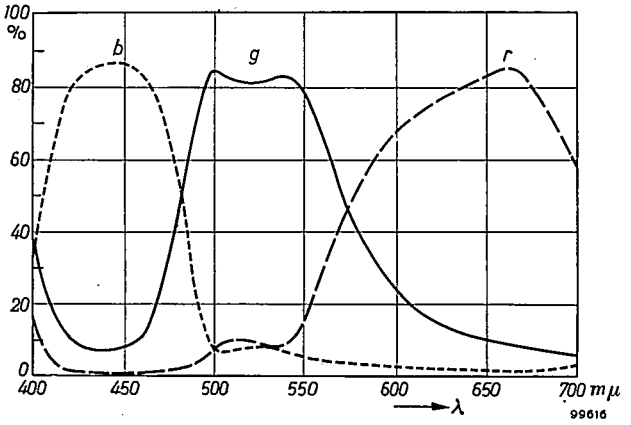


Fig. 9. Spectral characteristics of the cruciform arrangement of dichroic mirrors (*8* in fig. 4). As a function of wavelength λ the curve r represents the product D_bR_r, curve g the product D_rD_b and curve b the product D_rR_b, where D denotes the transmissivity and R the reflectivity. The suffixes r and b represent red and blue, respectively. The curves relate to an angle of incidence of 45°.

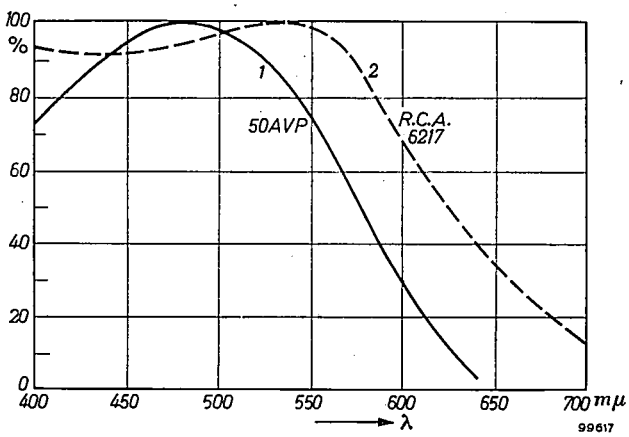


Fig. 10. Relative spectral sensitivity of two types of photomultiplier tube: Philips 50 AVP and R.C.A. 6217. Both curves are plotted relative to their peak sensitivity (= 100%). In the red region of the spectrum the type 6217 tube has a greater sensitivity, both relatively and absolutely.

This situation is worsened if, in order to bring the currents in the three channels to the same level, the 50 AVP tube is made to deliver a higher secondary-emission amplification by increasing the potential differences between the dynodes. Each multiplier tube adds to the signal current proper a noise component which is partly due to the presence of that current and partly to the "dark" current of the tube, which flows even when no light is incident on the photocathode. The dark current springs from several causes, such as thermionic emission from the photocathode, field emission from components of the tube, undesired photo-emission which occurs when fast electrons in the tube excite luminous phenomena that are "visible" to the photocathode, and finally leakage currents. Nearly all these unwanted effects become rapidly more serious if the voltage between the electrodes is raised in order to obtain a higher amplification. For this reason it was necessary in the case of the "red" channel to look around for a photomultiplier tube possessing a better spectral sensitivity in the red part of the spectrum than the type 50 AVP. The choice fell on the type 6217 tube made by the Radio Corporation of America; curve *2* in fig. 10 shows the spectral sensitivity of this tube (both curves are drawn with their peak sensitivity put at 100%).

The use of photomultiplier tubes specially sensitive to red has one disadvantage: the thermionic emission of the photocathode — which is part of the total dark current and has the character of noise — is troublesome even at room temperature. In spite of this, a 6217 tube is used in the "green" as well as in the "red" channel. The reason is that, if a 50 AVP were used, its decreasing sensitivity in the region from 540 to 610 mμ, in conjunction with the decreasing emission of the phosphor in the same region, would make a correction filter necessary having a relatively high absorption around the maximum of $\bar{G}(\lambda)$; see fig. 8. If a more red-sensitive tube is taken, such as the type 6217, it is sufficient to use a fairly light yellow-transmitting filter. Since the phosphor emits its highest radiant power in the green part of the spectrum, the effect of noise is negligible.

Having decided on the types of photomultiplier tube to be used, we can calculate the spectral transmission characteristics of the necessary correction filters. The next step is to select from the filter manufacturers' catalogues the type of filter that best approaches the required transmission.

Fig. 11 shows the spectral sensitivities calculated for our slide scanners.

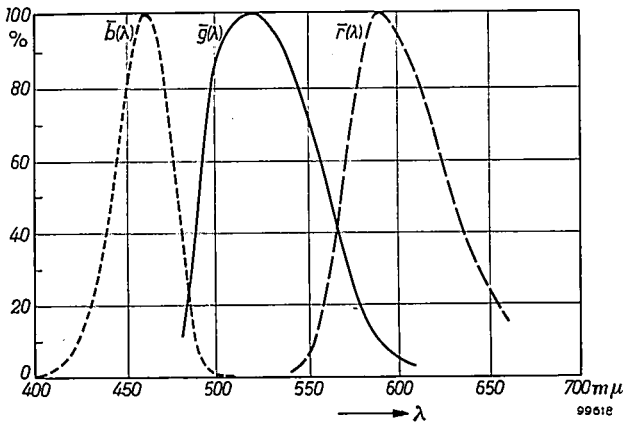


Fig. 11. Spectral sensitivity of the three colour channels (normalized to 100%) of the slide scanner, calculated from the known spectral characteristics of the phosphor, the dichroic mirrors, the correction filters and the photomultiplier tubes.

The electronic circuits

The deflection circuits and the extra-high tension generator of a flying-spot scanner for colour transparencies do not differ essentially from those published earlier. Since a flying-spot scanner is a signal source, higher demands are made on the linearity of the horizontal and vertical deflection than in picture tubes. The output voltage of the extra-high-tension generator is required to be extremely stable with respect to mains-voltage fluctuations, temperature variations and load variations. If the voltage is constant, the sharpness of the luminous spot on the cathode-ray tube need seldom be adjusted. This is particularly desirable in laboratory experiments, where different values of beam current may be used in fairly rapid succession. It may also be necessary now and then to vary the amplitude of the deflections; for this reason the extra high tension is taken from a separate generator and not, as is the practice in television receivers, derived from the horizontal deflection circuit.

The video amplifiers used in our flying-spot scanners for colour slides have a threefold function, viz.:

a) afterglow compensation,
b) gamma correction, and
c) distribution of the primary colour signals.

These functions are discussed in turn below.

a) *Afterglow compensation.* The phosphor on the screen of the scanning tube has the property of continuing to emit light for some time after excitation. Consequently the photomultiplier tubes receive light not only from the scanned picture element but also from picture elements previously scanned. Upon a sudden transition from dark to light or from light to dark in the slide, the picture signal therefore changes only gradually to the new value, resulting in a corresponding gradual transition in the received picture. This causes a certain unsharpness in the line scan of the picture, an effect also found if video amplifiers having too narrow a frequency band are used, i.e. the sharp transitions are blurred. Detailed quantitative considerations of afterglow are given in the article quoted under [6]). The circuit for afterglow compensation (*fig. 12*) is designed on the principles described in that article (p. 225). The basic idea is that the behaviour of the afterglow can be fairly accurately described by the sum of three exponential functions. In the cathode lead of the pentode in fig. 12 there are thus three *RC* networks whose time constants are made equal to those of the three exponential functions. One resistance and one capacitance are variable, in order to be able to make allowance for the aging of the phosphor.

Fig. 12 will be recognized as a circuit that selectively amplifies the higher frequencies (in that as the frequency rises the admittance of the capacitors bypassing the cathode resistors increases). This agrees with the described effect of the afterglow, namely that the system behaves like an amplifier with too narrow a frequency band; the band is widened by a circuit such as that in fig. 12.

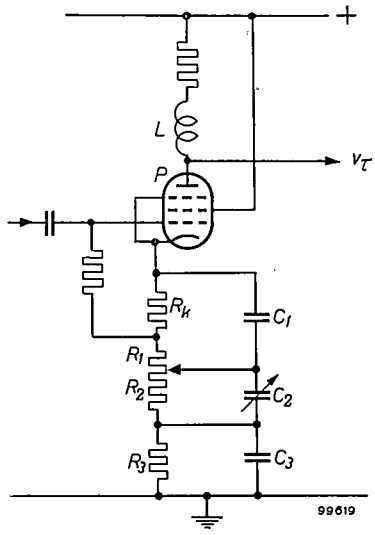


Fig. 12. Circuit for afterglow compensation. The sum of three exponential functions gives an approximation to the afterglow curve. The time constants $(R_k + R_1)C_1$, R_2C_2 and R_3C_3 of the RC networks in the cathode lead of the pentode P are made equal to the time constants of these three exponential functions. The magnitude of R_k ($\ll R_1$) determines the anode direct current. The ratio $R_1 : R_2$ is variable to allow for the properties of the phosphor changing with age. The inductance L in the anode circuit compensates for the loss in gain at high frequencies as a result of stray capacitances.

b) Gamma correction. The picture signal v_τ available following afterglow compensation is linearly proportional to the transmission of the slide. If this signal were directly used as a control signal for the picture tube, the picture produced would not give a true impression of the image on the slide. The reason is the non-linear relation between the luminous flux Φ and the control voltage V, which exists both in direct-viewing tubes and in projection tubes. As a rule the relation is $\Phi \propto V^\gamma$ (V measured from the cut-off point), where γ lies between 2.2 and 2.5. With suitable circuits, however, the signal v_τ can be made to produce a voltage $v \propto v_\tau^{1/\gamma}$, which is then suitable as the control signal, since $\Phi \propto v_\tau^{(1/\gamma)\gamma} = v_\tau$. In this way, then, we also linearize the last link in the transmission chain, namely the brightness of the picture tube.

The process by which a signal $v \propto v_\tau^{1/\gamma}$ is derived from the signal v_τ is called gamma correction. Many circuits have been devised for this purpose, and a feature common to all of them is that they contain one or more non-linear elements. The circuit used in our case is represented in *fig. 13a*. The triode T is arranged as a cathode follower. The variable resistance R is always large with respect to the resistance R_s, and R_s is large in relation to the resistance in the forward direction of the diode D. The resistance R is adjusted so that the D.C. potential of the point e is exactly equal to zero in the absence of a signal at g. A denotes a circuit functioning as a synchronous switch.

The synchronous-switch circuit A ensures that at point g the picture signal is zero during the flyback, i.e. when the scanning tube emits no light and the signal thus corresponds to black in the slide. The switch A is "closed" during the flyback or part of it; at times in between it is "open". (A circuit with a clamping diode can be imagined in place of A.)

As long as A is open the charge on the right plate of the coupling capacitor C cannot change; the potential of this plate thus varies together with that of the left plate, each scanning line, however, beginning from earth potential. We assume now that the input voltage v_τ has, say, a sawtooth waveform

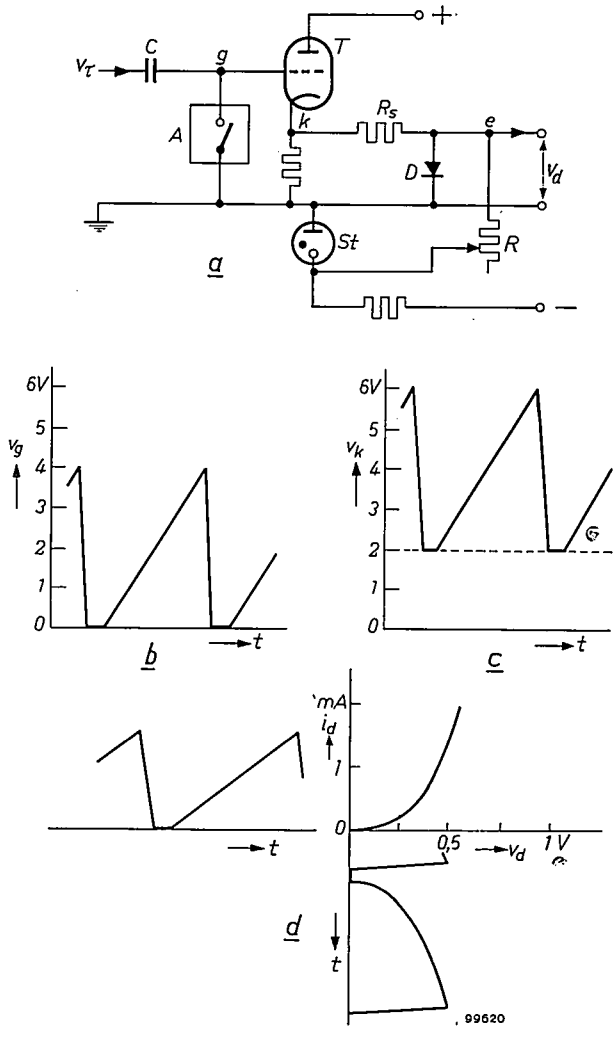


Fig. 13. *a*) Circuit for gamma correction. T triode, arranged as a cathode follower. C coupling capacitor. A is a circuit functioning as a synchronous switch, which suppresses the picture signal at point g during the flyback. D diode. St voltage reference tube. R resistor for reducing the D.C. potential at point e to zero, in the absence of a signal at g.
b) For the purposes of illustration it is assumed that the voltage v_g at point g in (*a*) has a sawtooth waveform.
c) The corresponding voltage v_k on the cathode.
d) A current i_d flows through the diode D. Owing to the curved characteristic of the diode, this current gives rise to an output voltage v_d having a curved sawtooth form, which compensates the opposite curvature of the characteristic $\Phi = f(V)$ of the picture tube.

(fig. 13*b*); through R_s and the diode the triode then sends a current i_d which is also of sawtooth shape (fig. 13*d*). Because of the curved characteristic of the diode the voltage v_d across the diode (likewise the output voltage) assumes a curved shape (fig. 13*d*). If we take a diode having a suitable characteristic we obtain $v_d \propto v_\tau^{0.4}$, which corresponds to $\gamma = 2.5$. For the three video amplifiers of a colour-slide scanner, three diodes are required that satisfy the above relation between 5 and 100% of the voltage range and in addition differ as little as possible in their individual characteristics.

The process described ensures that the dark parts are reproduced in the correct colour ratio, brighter than they would be without gamma correction. There are circumstances, however, in which an additional correction would be desirable, where some parts of the reproduced picture still appear too dark in relation to the bright parts. Such "hard" pictures may be obtained when contrasty slides are used. The contrast ratio of the slide, i.e. the ratio of the highest level of brightness to the lowest level of brightness, is then greater than that obtainable on the picture tube screen. The effect of this is that dark parts of the picture show too little contrast. With films and slides this is a frequent occurrence; since these are primarily intended for direct optical projection the difficulty does not arise here to the same extent, because an optical projector gives a higher contrast ratio than the picture tubes hitherto available for colour television. Sometimes it may be desirable for purely aesthetic reasons to reproduce dark parts with a relatively higher level of brightness.

The additional correction referred to can be obtained with a scanner for monochrome slides by reducing the exponent of the gamma correction (see article [6]) page 228). A glance at *fig. 14* shows that such a reduction is not permissible in the case of a scanner for colour slides. Suppose that we have selected $1/\gamma = 0.32$; then v_τ undergoes a correction $v_d' \propto v_d^{0.8}$ over and above the correction $v_d \propto v_\tau^{0.4}$. This can be seen in fig. 14. If a certain fragment of the image has a brightness level and a colour determined by the characteristic values R, G and B after the normal gamma correction, the additional correction would give this fragment a brightness and colour corresponding to the signals $R_{0.8}$, $G_{0.8}$ and $B_{0.8}$. This brightness level is indeed higher, as desired, but now the colour has changed too, since the "blue" signal has increased more than the "green" signal, and the latter more than the "red"; the colour balance has thus been destroyed.

The solution has been found in circuits that amplify the weaker picture signals more than the stronger ones, whilst preserving the existing ratio between the amplitudes of the "red", "green" and "blue"

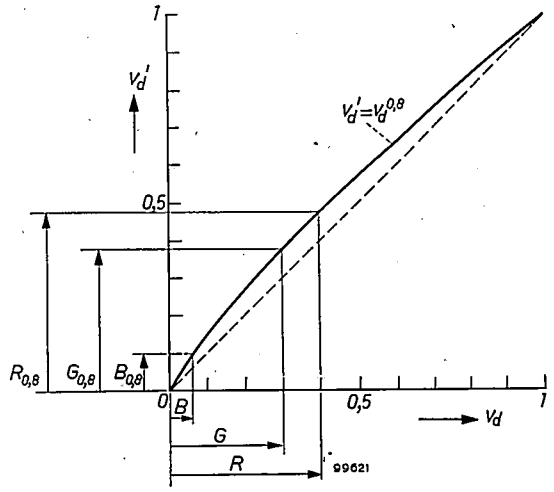


Fig. 14. If too hard colour pictures were corrected with the circuit in fig. 13*a*, they would not only become softer but also change colour, as illustrated here. R, G and B represent the three primary-colour signals (already gamma-corrected) which together constitute a patch in a colour picture which is desired to be softer. If each signal is passed through a circuit as in fig. 13*a*, having a characteristic $v_d' \propto v_d^{0.8}$, three additionally-corrected colour signals are obtained, $R_{0.8}$, $G_{0.8}$ and $B_{0.8}$, which are all greater in amplitude than R, G and B (a signal of value 1 would have remained unchanged). The relevant patch is thus made brighter. Whereas R and G have increased by 20 to 30%, B has almost doubled; the patch has thus changed in colour, which was not the intention.

signals. A circuit meeting these requirements and used successfully in the Philips laboratories has been described elsewhere [11]).

c) Distribution of the primary-colour signals. The three colour signals have to be routed to a test circuit and various monitors, and also to the various destinations required for experimental purposes. In order to make the various load circuits independent of each other, an output stage having a very low internal impedance is necessary. This brings us to the third and last function of the video amplifiers, i.e. the routing of the corrected signals.

Fig. 15 shows the circuit diagram of the distribution stage used in each of the three channels. As mentioned in article I, the three primary colour signals are commonly standardized so as to give them equal amplitudes for white (where absorption in the slide is independent of the wavelength of the light). For the output voltage of the preceding stage we selected a peak-to-peak value of 1 V. This, then, is the input voltage of the distribution stage in fig. 15. After being amplified by the double triode T_1-T_2, the signal drives a pentode P arranged as a

[11]) J. Kaashoek, Gradationsentzerrung im Farbfernsehen, Nachr.techn. Z. **11**, 515-518, 1958.

cathode follower. From the cathode of P negative feedback takes place via a resistor R_2 to the cathode of T_1. The voltage gain is just about equal to the ratio $(R_1 + R_2)/R_1$, where R_1 is the cathode resistance of T_1. Since $R_2 = R_1$, the potential at the point O is twice as high as the input voltage, i.e. 2 V peak-to-peak. Owing to the negative feedback the internal impedance is extremely low (< 1 ohm).

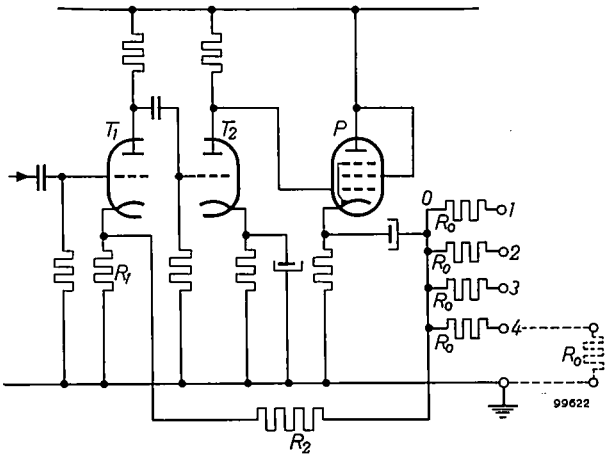


Fig. 15. Basic diagram of the distribution amplifier. T_1-T_2 double triode. P pentode as cathode follower. From the cathode of P feedback is effected via the resistance R_2 to the cathode of T_1 (with cathode resistance $R_1 = R_2$). *1*, *2*, *3*, *4* are output terminals to which are connected cables (having a characteristic impedance R_0 and terminated by R_0), leading to a test circuit, monitors and the consumer circuits. Input voltage 1 V peak-to-peak. Voltage at O: 2 V peak-to-peak. Voltage at the end of the cables: 1 V peak-to-peak.

The four output terminals *1* . . . *4* are each connected to point O via resistors R_0. Connected to the terminals are cables, each of characteristic impedance $R_0 = 135$ ohms, which are terminated at the input end by the resistances R_0, and must be terminated at the other end by the same resistance. Across each terminating resistor there thus appears a signal of 1 V peak-to-peak. The four cables lead respectively to

1) a second distribution amplifier, to which the various consumer circuits are connected,
2) a black-and-white monitor, used for checking the definition and geometry of the picture and to help in fault-finding,
3) a colour monitor [12]), for critically appraising the colours, and
4) a test circuit for measuring the output signals.

The advantage of this method compared with the use of, say, a cathode follower in each of the four channels is that the signals on the terminals *1* . . . *4* are completely identical. If no anomalies are found on the monitors or with the test circuit, one can be quite sure that correct signals are being distributed. Owing to the low internal impedance, a disturbance at the end of one of the cables (e.g. due to a mismatch) is not noticeable at the other places.

Finally, a few words about the test circuit referred to. The three primary-colour signals are displayed simultaneously side by side on a cathode-ray oscilloscope. The oscilloscope is provided with a D.C. restorer, which ensures that the parts of the oscillogram corresponding to "black" are always at the same level. A transparent screen fitted over the oscilloscope screen is provided with a horizontal line, previously traced through the peaks of a sinusoidal calibrating signal [13]) of 1 V peak-to-peak and having the same frequency as the horizontal deflection. When a slide is changed, the waveform of the primary-colour signals is observed on the oscilloscope, and the diaphragm of the objective lens is adjusted until the peak of the largest of the three signals just reaches the line; this signal then has an amplitude of 1 V peak-to-peak.

When the scanner is first switched on, a black-and-white slide is first inserted and then, by varying the supply voltage, the amplification of the photomultiplier tubes is regulated until the three primary-colour signals have the same amplitude.

Figs. 16 and *17* show two flying-spot scanners for colour slides, built on the principle described. One is a laboratory version, the other a version further developed by the Philips ELA Division.

Flying-spot scanner for coloured prints

Single-reflection arrangement

Fig. 18a and *b* shows the simplest arrangement of a flying-spot scanner for opaque matter, for black-and-white only. The raster of the scanning tube *1* is projected as an enlarged image on to the print *3* which is to be reproduced. Depending on the density (blackening) at the place where the image of the luminous spot happens to be, a greater or lesser part of the incident light is reflected; the remainder is absorbed in the paper. Of the reflected light the photocathode of the multiplier tube *4* picks up the part bounded by the cone *5*; this part generates the output current, which causes the picture signal to appear across the signal resistor. The signal is passed through a video amplifier identical with that used in a slide scanner. Here too, of course, afterglow compensation

[12]) See, for example, fig. 19 in the article under [2]).

[13]) The calibrating signal is generated by an oscillator the principle of which is described in: L. Ensing and H. J. J. van Eyndhoven, An oscillator with constant output voltage, Philips tech. Rev. 14, 304-312, 1952/53.

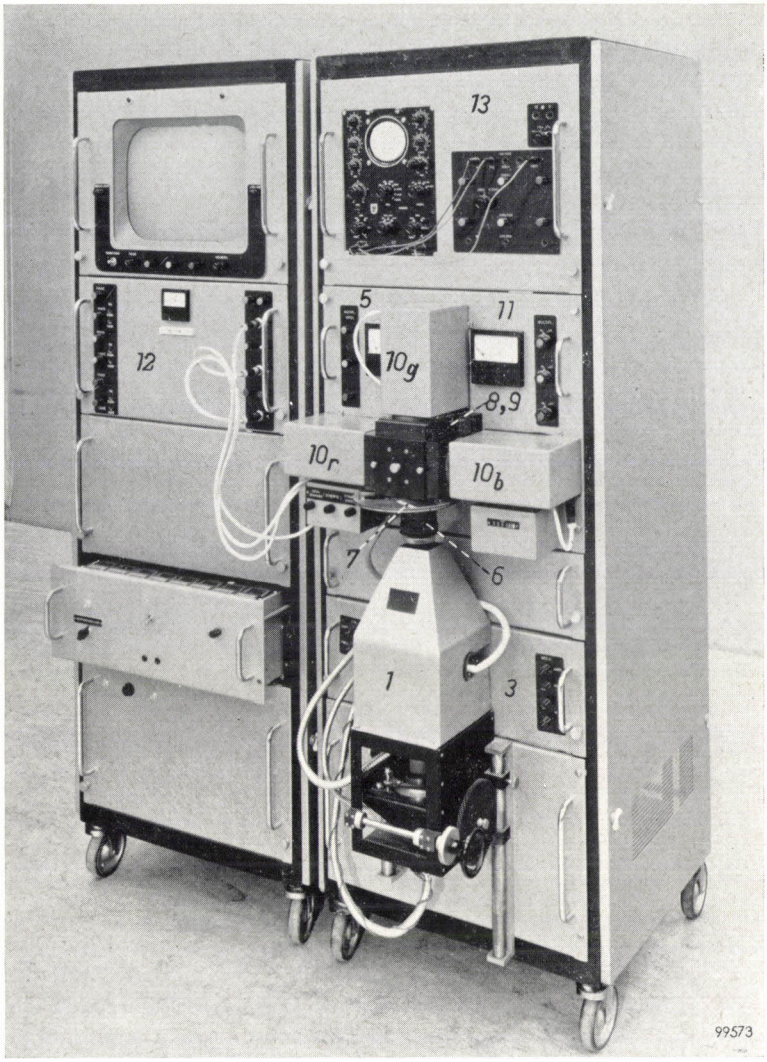


Fig. 16. Flying-spot scanner for colour slides, based on the principles described and built in the Philips Research Laboratories, Eindhoven. Meaning of figures as in fig. 4.

and gamma correction are equally necessary, to allow for the decay of the phosphor and the curved characteristic of the picture tube.

In the following we shall discuss some of the considerations underlying the design of a scanner for coloured prints.

If the print is a perfectly diffusing surface, the luminous intensity in any direction obeys Lambert's cosine law, irrespective of the direction from which the light strikes the surface. Let I_0 be the luminous intensity in the direction of the normal, then in a direction making an angle α with the normal the luminous intensity is $I_0 \cos \alpha$ (circular diagram, fig. 18*c*). If, on the other hand, the paper is glossy, the reflection is again diffuse except in directions near to the reflection angle equal to the angle of incidence. In these directions there is specular reflection, which is much stronger, and the diagram assumes the form of a lobe (fig. 18*d*), having a maximum and a width that depend on the nature of the glossy paper. Most photographs on art paper, as often used for colour reproductions, have reflecting surfaces of this kind, and so also have lacquered and plasticified types of paper as often used for advertisements and wrappings.

When the photomultiplier tube is positioned as shown in fig. 18*a*, the photocathode just remains clear of specularly reflected light (denoted by the dashed line). The luminous flux received by the photocathode from the middle and the edges of the

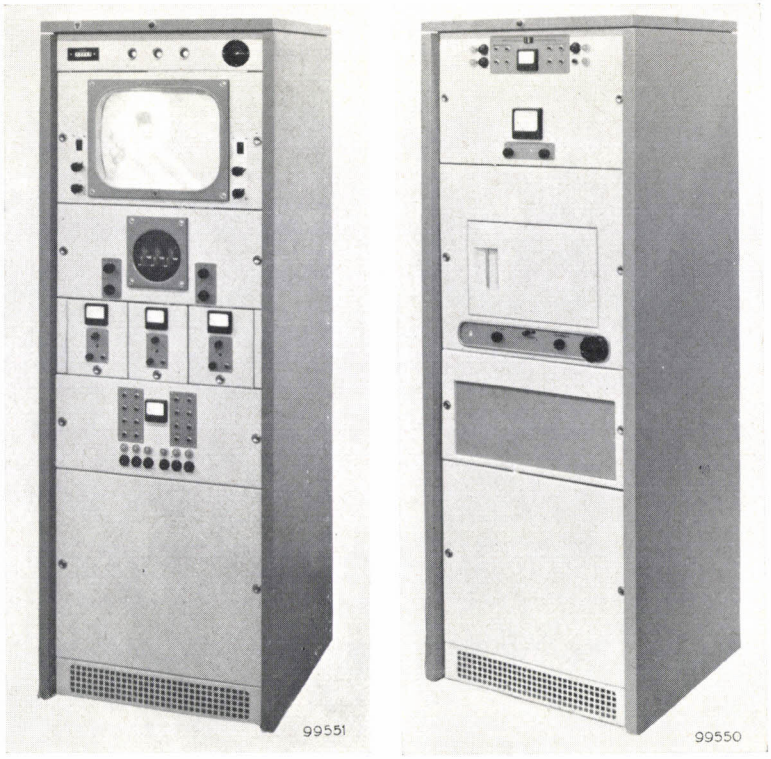


Fig. 17. The flying-spot scanner as in fig. 16, after further development under the direction of F. H. J. van der Poel in the ELA Division of Philips. The console on the left contains the monitor and the oscilloscope, that on the right the scanning tube, the optical system and the photomultiplier tubes.

paper, however, will differ. The angle δ can be so adjusted that most light will be received from the middle. The television picture will then show a vignetting effect to an extent that depends on the height of the multiplier tube above the paper and the distance from the optical axis. This effect, added to the unavoidable vignetting effect of the objective lens, may exceed the permissible limit if, in order to receive more light, the photomultiplier tube is positioned nearer to the print. In the case of colour television there is a further complication, in that three photomultiplier tubes have then to be arranged in such a way that the vignetting effect for all three is the same; otherwise, of course, colour defects would appear in the picture. All three photocathodes, then, should really have the same relative position. (This might be achieved by means of dichroic mirrors. These have not been used here, however, because of the light losses they cause and the constructional difficulties involved. In the arrangement such as in fig. 18*a* the light is separated into the three primary colours entirely by filters mounted in front of the photocathodes; we shall return to this point at the end of the article.)

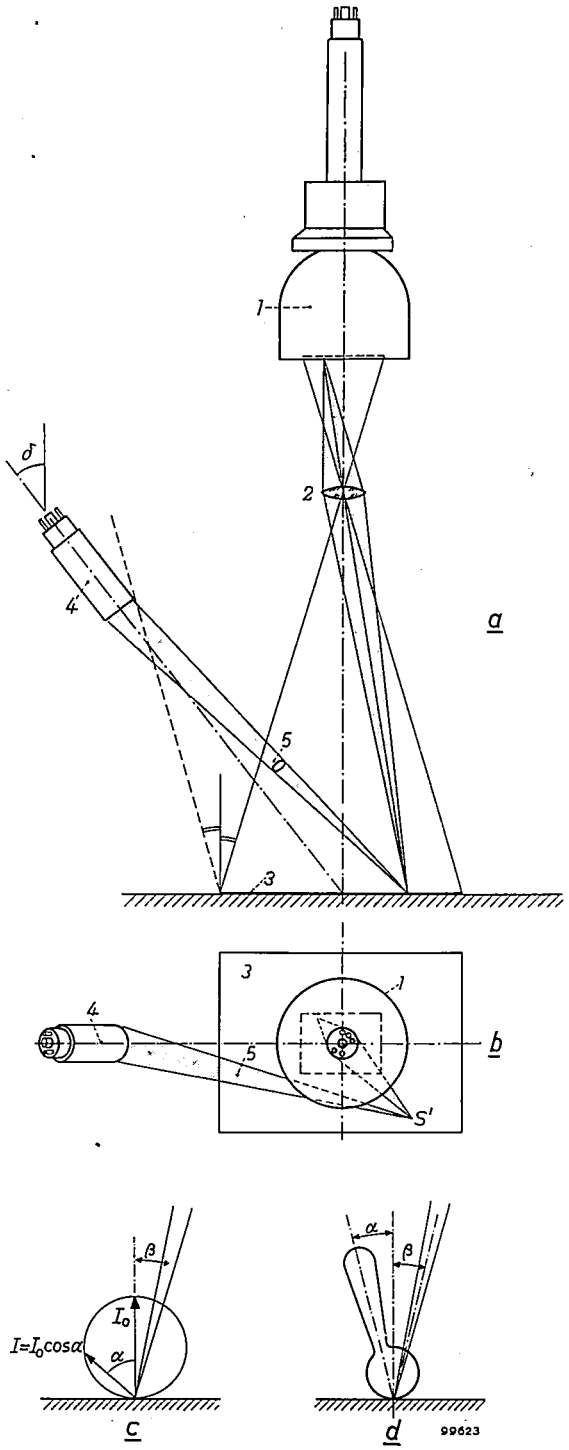


Fig. 18. *a*) Vertical section, *b*) horizontal projection, of a simple arrangement for a print scanner. *1* scanning tube. *2* objective lens which projects the raster on the scanning tube screen upon the print *3*. *4* photomultiplier tube. *5* cone within which the photocathode of *4* is seen by a scanned picture element *S* of the print.
c) Diffuse reflection: luminous intensity I in the direction α is proportional to cos α (Lambert's law), irrespective of the direction of incidence β.
d) Specular plus diffuse reflection from glossy paper: the luminous-intensity diagram shows a lobe in the direction $\alpha = -\beta$.

Another point of importance in a set-up as in fig. 18*a* is that careful precautions must be taken to prevent light other than that from the luminous spot from falling on to the print. "Alien" light in the first place produces an additional D.C. component in the output current; although the resultant direct-voltage component is easily removable from the actual picture signal, the additional current increases the noise, thus causing needless deterioration of the signal-to-noise ratio. The interfering light is usually artificial light from lamps operating on alternating current; this will therefore give rise to a second undesired effect in the picture signal, in the form of a hum component of twice the mains frequency. If the interfering light is relatively intense, it might in addition produce non-linear distortion phenomena.

The arrangement described below provides effective shielding against interfering light and considerably reduces vignetting effects.

Scanner using a photometer sphere

Fig. 19 illustrates a set-up in which the objective lens and the multiplier tubes are mounted in the wall of a hollow sphere provided with a diffusing white internal surface. The use of a photometer sphere offers an appreciable light gain [14]). The photo-

[14]) A print scanner using a photometer sphere for monochrome television is described in: H. Stier, P. Lindner and E. Kosche, Die Lichtpunktabtastung und die Übertragung episkopischer Bildvorlagen, Nachr.technik **5**, 537-541, 1955.

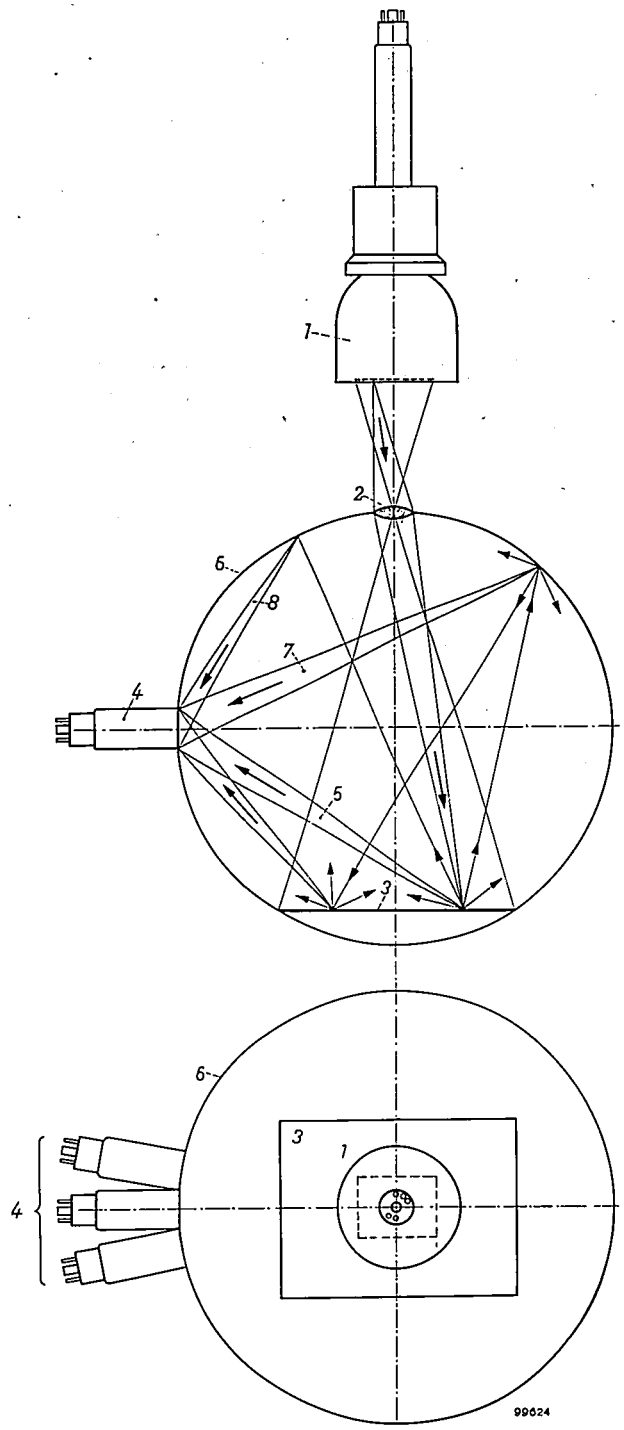


Fig. 19. Print scanner based on the use of a photometer sphere. *1* scanning tube. *2* objective lens. *3* print. *4* photomultiplier tubes. *5* cone of light directly incident on a photocathode. *7*, *8* cones of light reaching the photocathode after a single reflection from the wall of the photometer sphere *6*.

cathodes now receive not only the light reflected directly from the print (cone *5*) but also the light diffusely reflected once (cones *7* and *8*) or more than once from the wall of the sphere. The light that in fig. 18*a* was reflected outside the bounds of the cone *5* was entirely lost, whereas now a great deal of it reaches the cathodes. The gain provided by this integrating effect amounts to a factor F, given by the expression

$$F = \frac{1}{1 - A\varrho_w},$$

where ϱ_w is the reflection coefficient of the white inner wall of the sphere, and A is the part of the wall participating in the diffuse reflection in relation to the total surface area of the sphere (the objective lens, the photocathodes and the print itself are considered to make no contribution to the diffuse reflection).

In *fig. 20a*, S represents a diffusely radiating surface. In a direction at an angle α with the normal to S the luminance L_α is by definition the ratio of the luminous intensity I_α (in candelas) and the effective surface $S \cos \alpha$ (in m²), both seen from the direction α:

$$L_\alpha = \frac{I_\alpha}{S \cos \alpha} \text{ nit.}$$

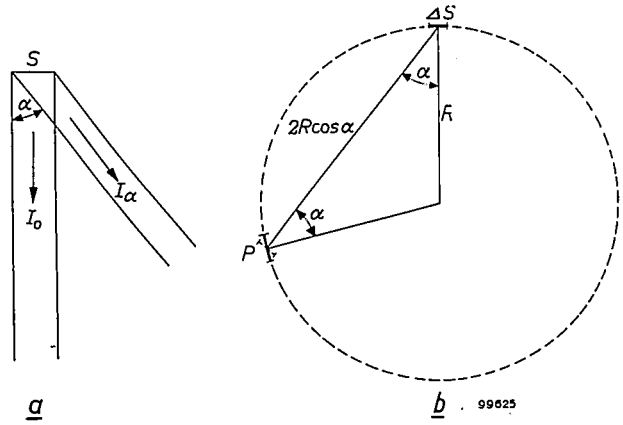


Fig. 20. *a*) In a direction α with respect to the normal, a diffusely radiating surface S that obeys Lambert's law has luminous intensity $I_\alpha = I_0 \cos \alpha$.
b) A surface ΔS, obeying Lambert's law, and a surface at P form part of an imaginary sphere of radius R. It is shown in the text that the illumination of the surface at P is independent of the location of P on the sphere.

If the luminance is the same in all directions, it follows that

$$I_\alpha = L\, S \cos \alpha = I_0 \cos \alpha,$$

where $LS = I_0$. $I_\alpha = I_0 \cos \alpha$ is simply Lambert's law. The luminous flux Φ emitted by the surface S, where Lambert's law is obeyed, is calculated to be:

$$\Phi = \pi I_0 \text{ lumens.}$$

In fig. 20*b* the surface ΔS represents a diffuse light source of this kind, having a luminance L. The luminous intensity in the direction α is then $L\Delta S \cos \alpha$. We now calculate the illumination E_p on a surface through the point P, which surface, like ΔS, forms part of an imaginary sphere of radius R. For E_p we can write:

$$E_p = \frac{L\Delta S \cos \alpha}{(2R \cos \alpha)^2} \cos \alpha = \frac{L\, \Delta S}{4R^2}.$$

The illumination is thus independent of the place where P is situated on the wall of the (imaginary) sphere.

For the emitted luminous flux Φ we can write $\Phi = \pi I_0 = \pi L \Delta S$, and therefore $E_p = \Phi/4\pi R^2$.

In the arrangement shown in fig. 19 let Φ be the luminous flux entering the photometer sphere through the objective lens, and let ϱ_p be the reflection coefficient of the paper at the position of the luminous spot at a given moment. Then $\varrho_p\Phi$ is the luminous flux reflected by the print at that moment. Assuming that the print obeys Lambert's law, the wall of the sphere receives a uniform illumination of intensity $E_1 = \varrho_p\Phi/4\pi R^2$ (disregarding for the present the reflections from the sphere wall). Certain parts of the sphere wall, however, are not painted white, namely the objective lens, the photocathodes and the print itself. We assume that these parts do not reflect at all, and take this into account by the factor $A\,(< 1) =$ white surface area/total sphere surface area. If ϱ_w be the reflection coefficient of the white part, and $\varrho_p\Phi$ the luminous flux emanating from the print, the luminous flux reflected by the wall for the first time is $A\varrho_w\varrho_p\Phi$. This returns into the sphere and increases the illumination by an amount $\Delta E = A\varrho_w\varrho_p\Phi/4\pi R^2$. This process is repeated an infinite number of times. The total illumination finally becomes:

$$E = \lim_{n=\infty} \frac{\varrho_p\Phi}{4\pi R^2}\,(1 + A\varrho_w + A^2\varrho_w^2 + \ldots + A^n\varrho_w^n) = \frac{\varrho_p\Phi}{4\pi R^2}\,\frac{1}{1 - A\varrho_w}. \quad \ldots\ldots \quad (1)$$

Without the photometer sphere the illumination would be:

$$E_1 = \frac{\varrho_p\Phi}{4\pi R^2}. \quad \ldots\ldots\ldots \quad (2)$$

By using the photometer sphere we thus gain a factor

$$F = \frac{1}{1 - A\varrho_w} \quad \ldots\ldots\ldots \quad (3)$$

in the illumination of the sphere wall, and hence in the luminous flux incident on the photocathodes.

In *fig. 21* the gain factor F is plotted as a function of ϱ_w, for various values of A. It may be noted at this stage that the sphere can be given a reflection coefficient ϱ_w of 0.95 (we shall return to this later). It can be seen from fig. 21 that with $A = 0.85$, 0.90 and 0.95 (i.e. with respectively 15, 10 and 5% of the sphere wall non-reflecting) the gain factor is respectively about 5, 7 and 10. The choice of A will also be dealt with later.

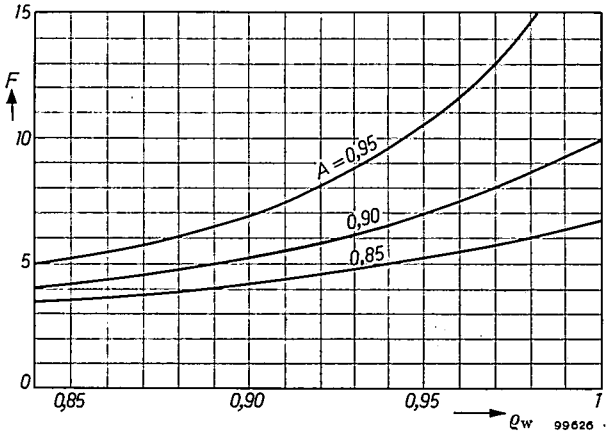


Fig. 21. The gain $F = (1 - A\varrho_w)^{-1}$ in luminous flux on the photocathode when using a photometer sphere, as a function of ϱ_w for various values of A.

Apart from the appreciable gain in light, important advantages of the sphere are that it shields the system against interfering light and greatly reduces the vignetting effects mentioned above. A further advantage is that the sphere attenuates the colour defects caused if the scanner is used for displaying small three-diemensional objects. In a scanner without such a sphere the objects are seen by the photomultiplier tubes from slightly different directions, with the result that they appear in the picture with coloured edges. The sphere largely eliminates this effect.

Comparison of luminous fluxes on the photocathodes in scanners for slides and for prints

In order to compare the luminous flux on the photocathodes in scanners for slides with that on the photocathodes in print scanners we first calculate, with the aid of *fig. 22*, the luminous flux

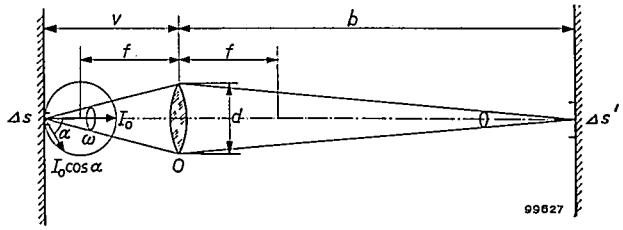


Fig. 22. Image formed of the luminous spot (surface area Δs) on the object to be reproduced (slide or print) by means of an objective lens O of diameter d and focal length f. The object distance is v, the image distance b.

Φ' which, at a given luminance L of the spot, is emitted from the objective lens. We assume that the phosphor obeys Lambert's law, i.e. that the luminous intensity of the spot at an angle α with the normal is $I_0 \cos\alpha$; the solid angle ω subtended at the objective by the spot is assumed to be so small, however, that we may put $\cos\alpha = 1$. Within the solid angle ω the luminous flux radiated is then $\Phi = \omega I_0 = \omega L \Delta s$, where Δs is the surface area of the spot. Let the transmission of the objective be τ_0; the luminous flux Φ' radiating from the objective is then

$$\Phi' = \tau_0\,\omega\,L\,\Delta s. \quad \ldots\ldots \quad (4)$$

To simplify the comparison we use the relations $\omega = \frac{1}{4}\pi d^2/v^2$, $d = f/c$ and $v = fN/(1 + N)$ to reduce equation (4) to

$$\Phi' = \frac{\pi\,\tau_0\,L\,\Delta s}{4c^2}\left(\frac{N}{1+N}\right)^2, \quad \ldots\ldots \quad (5)$$

where $c = f/d$ is the aperture of the objective, i.e. the ratio of its focal length to its diameter, and N is the magnification $= |b/v|$ (b = image distance, v = object distance; see fig. 22).

a) Slide scanner. The luminous flux Φ' falls on the colour slide and the light passes successively (fig. 4) through the slide, the first condenser lens, the dichroic mirrors, the second condenser lens and the correction filter. The luminous flux Φ_s' finally incident on the photocathodes is

$$\Phi_s' = \tau_s \tau_{os} \tau_{ms} \tau_{fs} \times \tfrac{1}{4}\pi L \Delta s \left(\frac{N_s}{1+N_s}\right)^2 \times \frac{1}{{c_s}^2}, \quad (6)$$

where

τ_s is the spectral transmission of the colour slide,
τ_{os} is the transmission of the objective lens,
τ_{ms} is the spectral transmission of the dichroic mirrors and the condenser lenses together, and
τ_{fs} is the spectral transmission of the correction filter.

(The suffix s denotes the case of the slides.)

b) Print scanner (with photometer sphere). Here the luminous flux Φ' is reflected from the print, and the photocathodes pick up the reflected light, in part directly and in part after single or multiple reflections from the wall of the sphere. The luminous flux incident on the photocathodes is equal to the illumination E of the sphere wall times the effective surface of the photocathode (of radius r) times the spectral transmission of the filter. Introducing the suffix p to denote the case of the print scanner, we thus have $\Phi_p' = E \times \pi r^2 \times \tau_{fp}$, from which, in conjunction with (1) and (3), we obtain:

$$\Phi_p' = \varrho_p \tau_{op} \tau_{fp} \frac{r^2}{4R^2} F \times \tfrac{1}{4}\pi L \Delta s \left(\frac{N_p}{1+N_p}\right)^2 \times \frac{1}{{c_p}^2}. \quad (7)$$

In order to compare Φ_s' and Φ_p', i.e. (6) and (7), we make the following assumptions:

1) The raster on the scanning tube is the same size in both cases and the same beam current is used, so that $\frac{1}{4}\pi L \Delta s$ is equal in both cases.
2) $\tau_{os} = \tau_{op}$ (the objective lenses have the same transmission).
3) $\tau_s = \varrho_p$ (the spectral transmission of the transparency is equal to the spectral reflection coefficient of the print).
4) $\tau_{ms} \times \tau_{fs} = \frac{1}{2}\tau_{fp}$ (mirrors and filters of the slide scanner together have the same spectral transmission as the filter in the print scanner, and the transmission of glass bodies, such as the condenser lenses and the bases of the dichroic layers, is assumed to be 0.5).

With these assumptions we find from (6) and (7):

$$\frac{\Phi_s'}{\Phi_p'} = \frac{2}{F}\left(\frac{R}{r}\frac{c_p}{c_s}\frac{N_s}{1+N_s}\frac{1+N_p}{N_p}\right)^2. \quad (8)$$

For the slide scanner described we have: $N_s = 0.367$ and $c_s = 2.8$ (this lens aperture is the one most commonly used). For the print scanner, described below, we have: $F = 6.9$ (obtained from $A = 0.90$ and $\varrho_w = 0.95$), $N_p = 3$, $c_p = 2.5$, $R = 20.2$ cm and $r = 2.3$ cm (in the case of the R.C.A. tube type 6217). These values inserted in (8) yield: $\Phi_s' = 2.3\ \Phi_p'$. If the assumptions made are valid, the photocathodes in the print scanner receive a luminous flux which is 2.3× less than that in the case of the slide scanner. In order to obtain voltages of the same amplitude at the output of the video amplifiers it is therefore necessary to increase the amplification of the photomultiplier tubes by a factor of 2.3, which will give rise to a slight decrease in the signal-to-noise ratio.

Of course, one might also try to modify the above assumptions in such a way as to improve the calculated ratio of Φ_s' to Φ_p', possibly even to make Φ_p' larger than Φ_s'. This brings us to the question of the design and dimensioning of the print scanner.

Design and dimensioning of the print scanner

The first point of importance is the size of the prints to be reproduced. It was decided that the maximum size of print to be scanned should correspond to the printed part of a quarto format, i.e. 18×24 cm. Since the raster on the scanning tube is usually 6×8 cm, the magnification has to be $N_p = 3$.

Another design parameter is the radius R of the sphere, and in deciding on this it is necessary to take the factor A into account. It is desirable that A should be as close as possible to 1, in the first place because, according to (3), the gain factor obtained from the sphere is then greatest. Another reason is that, if A is small, there is a risk of colour poisoning. In defining A we assumed that the prints (and also the objective lens and the photocathodes) had no part at all in the reflections. In reality, of course, there are reflections from the print, and their influence on the colour rendering is evident if we consider the extreme case of a print which is entirely red except for a small green patch. The "green" signal produced when this patch is scanned will be weaker than it would be if the rest of the print were also green. The effect diminishes the smaller is the surface area of the print in relation to that of the sphere, that is the higher the value of A. Since the joint surface area of objective lens, photocathode and print is more or less fixed, A can only be raised by increasing the radius R. It appears

from (8), however, that the ratio of the luminous fluxes $\Phi_p' : \Phi_s'$ is proportional to R^2, so that increasing R would be greatly to the disadvantage of the print scanner compared with the slide scanner as regards the luminous flux on the photocathodes. By way of compromise we decided on $A = 0.9$, which, with the given dimensions of print and photocathodes (those of the objective lens can only be estimated) leads to $R = 20$ cm.

If the objective is fitted in the sphere wall (fig. 19) the focal length f, now that R and the magnification N_p are fixed, is also determined and is, in this case, 9.25 cm. Lenses of this focal length and of f-number $c = 2.5$ are commercially available.

Finally, a few words about the way in which the inside of the sphere can be made reflective. From the optical point of view a vapour-deposited coating of magnesium oxide approaches the ideal, in that it obeys Lambert's law and has a reflection coefficient of 0.98. Its adhesion, however, leaves much to be desired. The article quoted under [14]) refers to a mixture of ground MgO with a solution of gelatin in sufficient water to produce a spray-paint; the reflection coefficient mentioned is 0.95. Experiments at Philips have resulted in a similar paint, but using lithopone (a mixture of ZnS and $BaSO_4$) instead of MgO; this showed somewhat better adhesion and its reflection coefficient is 0.955. With $A = 0.9$ the gain factor F is 7 (fig. 21). (For photometric purposes it is more important to have a durable paint than to achieve the utmost in light economy; the paint commonly used in integrating photometers has a reflection coefficient of 0.90, which would give an F of only 5.)

Some constructional details of the print scanner

On the lines described above an experimental print-scanner has been designed, the optical components of which appear in *fig. 23*. A cross-section of the scanner is shown in *fig. 24*. The first thing one notices is that the sphere has been replaced by a cube with cut-off corners. This is assembled from a frame fitted with rectangular and triangular cover plates, which are readily removable if a fresh coating of paint is necessary. The three photomultiplier tubes are mounted in a triangular arrangement in a box which, for the purposes of experiment, can be secured at different places. As indicated by a dashed line, none of the photocathodes can receive any light reflected specularly from flat glossy prints.

The second point to be noted is that the scanning tube is positioned horizontally. The vertical position shown in fig. 19, with the screen pointing downwards, is undesirable because of the danger of damage being caused to the phosphor layer by loose particles in the tube. A plane mirror at an angle of 45° reflects the light downwards, enabling the print to be placed in a horizontal position.

It can further be seen that the objective lens is not mounted in the wall of the cube but between the scanning tube and the mirror (the focal length must accordingly be greater than mentioned above). Both the tube and the objective lens can be slid along the frame on which they are mounted, thereby making it possible to "enlarge" parts of 18×24 cm prints or smaller prints. Fig. 24*a* shows the situation for scanning a print of 18×24 cm, fig. 24*b* the case for a print of only 11×14.5 cm. This "optical enlargement" can be supplemented by "electrical enlargement", that is to say by reducing the size of the raster on the tube window. If this is done over a long period, however, there is a risk of a burn mark appearing on the phosphor, which will be visible in the picture when reverting to a raster of normal dimensions.

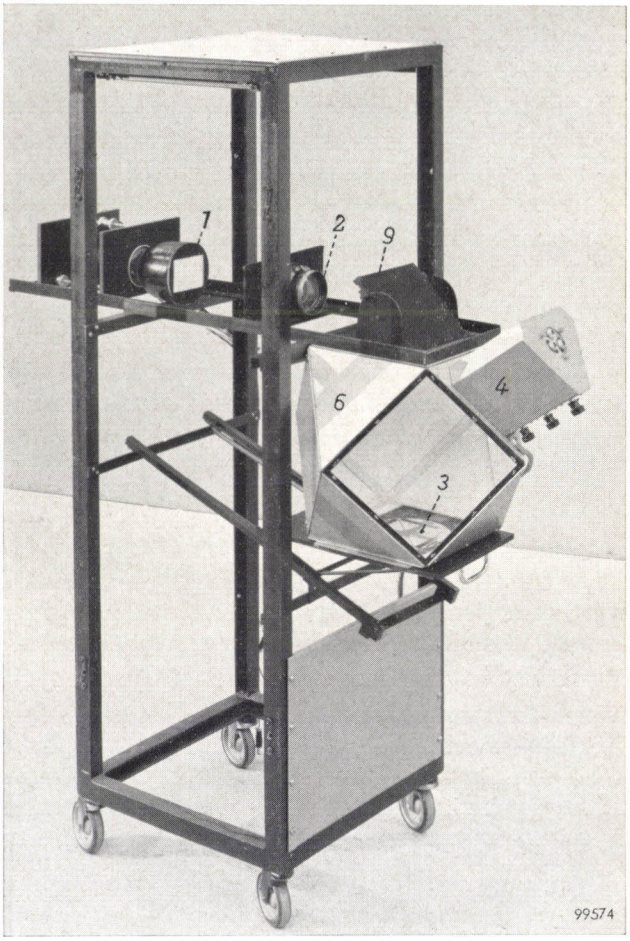
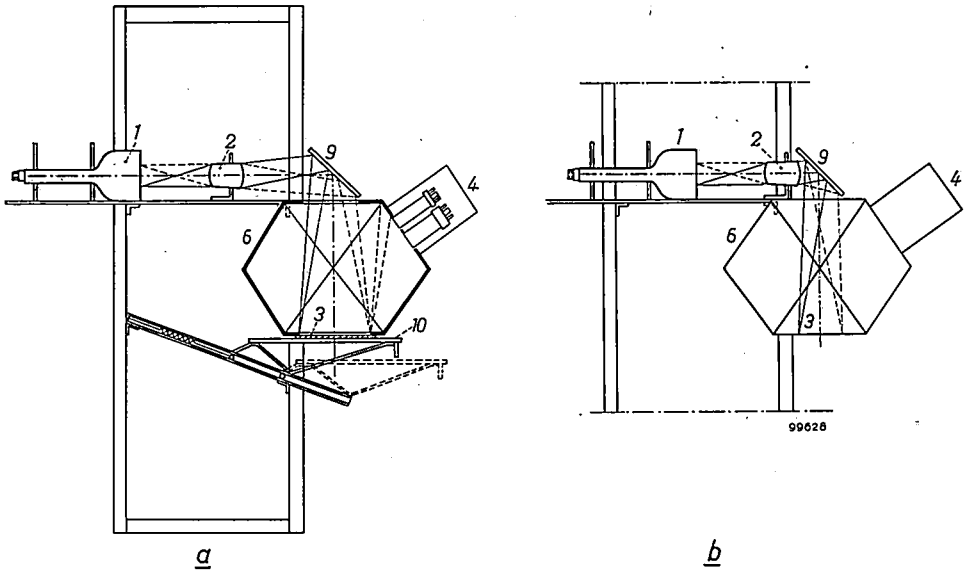


Fig. 23. Experimental print-scanner for colour television, built in Philips Research Laboratories, Eindhoven, photographed in the absence of the electrical sub-units to make the optical system clearer. *1* scanning tube. *2* objective lens. *3* print (here a test plate). *4* housing containing the three photomultiplier tubes. *6* "photometer sphere", here a cube with cut-off corners (one side plate has been removed to show the interior). *9* plane mirror at an angle of 45°.

As may be inferred from formula (7), enlargement will have the effect of decreasing the luminous flux on the photocathodes, and hence the output signal: in "optical enlargement" the magnification N_p of the optical system is actually *reduced*, and in electrical enlargement Δs is reduced. The loss of light can be considerably diminished, however, by covering the non-reproduced part of the print with a white surround; this then increases A, and moreover lessens the risk of colour poisoning.



a b

Fig. 24. *a*) Cross-section of the print scanner shown in fig. 23. The figures *1* . . . *9* have the same meaning as in fig. 23. *10* movable platform with the print at position *3*. Scanning tube and objective lens are in position for prints of 18 × 24 cm.
b) Scanning tube and objective lens in position for prints of 11 × 14.5 cm (maximum optical enlargement).

The possibility of enlargement enhances the value of the scanner for use in colour-television broadcasts, since it makes it possible to transmit all manner of coloured prints without faults of register from the signal source.

The spectral sensitivity of the three channels in the print scanner should follow the curves in fig. 8, as in the case of the slide scanner. With the latter this is achieved mainly by the use of dichroic mirrors, additional correction being provided by filters in front of the photocathodes. In the print scanner, however, these mirrors are lacking and the filters are solely responsible for separating the light into the primary colours. It is not an easy matter to find filters capable of meeting the requirements with sufficient accuracy. The solution adopted was to use several different filters, placed side by side over the photomultiplier. If, for example, each half of the photocathode is covered with a different filter, the net spectral transmission is the arithmetical mean of the transmissions of the individual filters. This artifice, which is by no means new, has produced good results.

Summary. In development work on colour television it is desirable that the source of the primary-colour signals should be completely free from errors of register. Colour-television cameras do not meet this requirement, but flying-spot scanners do, and for this reason they have been widely used in colour-television development. Two types have been evolved in the Philips Research Laboratories: one for colour slides and one for opaque matter (colour prints, paintings, coloured drawings, etc. and also small objects.) In both cases the flying spot is produced on the screen of a special cathode-ray tube whose phosphor shows maximum emission at a wavelength of 505 mμ.

In the slide scanner the flying spot is imaged on the transparency by an objective lens, and the objective aperture is imaged on the photocathodes of three photomultiplier tubes (one Philips type 50 AVP, two R.C.A. types 6217) by condenser lenses via dichroic mirrors and correction filters. The video channel of each of the three primary-colour channels has an afterglow compensation section, a gamma correction section and an output amplifier. The latter distributes the colour signals via cables to a test circuit, two monitors and the consumer circuits.

The scanner for colour prints is based on the principle of the integrating-sphere photometer. This provides a light gain of about a factor 7, and also shields the system against extraneous light and reduces vignetting effects. The design and various constructional details of such an experimental print-scanner are discussed.

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 Laboratories, Eindhoven, Netherlands.

2684: F. C. de Ronde: Schwinger's variational principle applied to the calculation of the radiation resistance and radiation reactance of a linear antenna in a waveguide of rectangular cross-section (Commun. Congrès int. Circuits et Antennes Hyperfréquences, Paris 21-26 Oct. 1957, part 1 (suppl. Onde électr. **38**, No. 376 bis), pp. 95-98, 1958).

The radiation impedance of a linear antenna in a waveguide of rectangular cross-section has been calculated on the basis of an assumed sinusoidal antenna current distribution. Because of the rather great discrepancy between theory and experiment a more powerful method is applied to calculate the radiation impedance, viz. Schwinger's variational principle. This new application, even when based on only a two-term Fourier expansion of the current distribution, gives better results than the assumed sinusoidal current distribution.

2685: F. C. de Ronde: A simple component for impedance measurements at cm and mm waves: the direct-reading variable impedance (Commun. Congrès int. Circuits et Antennes Hyperfréquences, Paris 21-26 Oct. 1957, part 1 (suppl. Onde électr. **38**, No. 376 bis), pp. 294-295, 1958).

At cm wavelengths standing-wave indicators are used for impedance measurements. However, it is not simple to make them for mm waves. Impedance measurements can also be done in a bridge. For that reason a variable impedance has been developed which, in combination with a hybrid tee, can be used for impedance measurements. This hybrid tee only has to be symmetrical, so it is not difficult to construct. The modulus and argument of the reflection coefficient of the variable impedance can be adjusted independently of each other. Variable impedances for 3 cm, 8 and 4 mm have been realized. The precision of the modulus of the reflection coefficient is a few percent.

2686: G. Ahsmann and H. J. Oskam: Impedance and recovery time of glow discharges in mixtures of rare gases (J. appl. Phys. **29**, 1768, 1958, No. 12).

Measurements of the impedance of a series of inert gas discharges (13.7 mm Hg, 3.5 mA) by means of a Wagner A.C. bridge, at frequencies of 100 to 15 000 c/s. From these measurements the characteristic quantities, viz. the self-inductance and the differential resistance, are calculated. The results agree well with a theory given by C. van Geel. Additions of a few percent of another inert gas, of lower ionization potential, cause a considerable drop in the modulus of the impedance. An explanation of this effect is given. The influence of the added inert gas on the recovery time will be dealt with in a later publication.

2687: B. Verkerk: Autoradiographic study of the influence of chlorine on semi-conducting alumina layers (Radioisotopes in scientific research, Proc. int. Conf. held in Paris, Sept. 1957, under the auspices of the UNESCO, Pergamon Press, London, New-York and Paris 1958, edited by R. C. Extermann, Vol. 1, pp. 503-515).

The presence of very small quantities of chlorine in the electrolyte has a disastrous effect on the insulating properties of anodically formed alumina layers. An autoradiographic method was used to study the presence and distribution of absorbed chlorine on alumina layers formed in a boric acid bath, to obtain information on the mechanism of this attack. A proposed mechanism attributing the electronic conductance of degenerated layers to a partial displacement of the oxygen in the oxide by chlorine with a corresponding change in the valency of part of the aluminium did not seem to hold in the light of the results obtained. The distributions found pointed toward the presence of imperfections in the oxide layers due to segregations of impurities at isolated places of the aluminium and along grain boundaries of recrystallized samples. Identical distributions were found for some absorbed cations. From this and the work of French investigators it is concluded that the corrosion by chlorine of anodized aluminium is due only to impurities in the metal.

2688: Dutch translation of **2690**.

2689: French translation of **2690**.

2690: H. B. G. Casimir: Certainty in the exact sciences (Koninklijke Nederlandse Akademie van Wetenschappen 1808-1958, N.V. Noord-

Hollandsche Uitgeversmaatschappij, Amsterdam 1958, pp. 243-251).

Philosophic considerations on the certainty that may be attributed to physical theories. Examples from other fields of human endeavour and from the historical development of physical theories are used to illustrate the theme. Even in physics certainty is not absolute: it is limited by the particular assumptions associated with a theory or by the circumstances in which observations are made.

2691: T. Kralt: Ionylamines, III. Catalytic hydrogenation of ionones in the presence of ammonia or amines and the determination of the position of the $>C=C<$ bond in dihydro-α-, dihydro-β- and tetrahydro-ψ-ionylamines (Rec. Trav. chim. Pays-Bas **77**, 990-1003, 1958, No. 11).

The course of the catalytic hydrogenation of ionones in the presence of ammonia or amines has been investigated. In this way it has been possible to determine the position of the remaining $>C=C<$ bond in the dihydro-α-, dihydro-β- and tetrahydro-ψ-ionylamines described previously.

2692: J. H. Uhlenbroek and J. D. Bijloo: Investigations on nematicides, I. Isolation and structure of a nematicidal principle occurring in Tagetes roots (Rec. Trav. chim. Pays-Bas **77**, 1004-1009, 1958, No. 11).

Some highly nematicidal polythienyls have been isolated from the roots of Tagetes plants. One of the compounds was shown to be α-terthienyl.

2693: J. P. L. Bots: Investigations on sterols, XI. Preparation of 13α-androstene derivatives (Rec. Trav. chim. Pays-Bas **77**, 1010-1017, 1958, No. 11).

By irradiation of androst-5-en-3β-ol-17-one with U.V. light ($\lambda > 280$ mμ) 13α-androst-5-en-3β-ol-17-one was obtained. This substance was converted into a mixture of the C_{17} epimers of 13α-androst-4-en-17-ol-3-one, both of which were obtained in a pure state.

2694: F. Brücke, G. Hertting, H. D. Moed and J. van Dijk: Über eine neue Adrenalincarbonsäure und einige ihrer Derivate (Biochem. Pharmacol. **1**, 221-231, 1958). (On a new adrenalin carboxylic acid and some of its derivatives; in German.)

Synthesis and pharmacological properties of a new carboxylic acid of epinephrine and of some derivatives are reported. The carboxylic acid produced qualitatively the same effects on the blood pressure and on the denervated nictitating membrane of the cat as epinephrine, the relative potency being 0.0003. The effects on the blood pressure could be reversed by previous administration of adrenalytics. The authors believe that the effects on the blood pressure and the nictitating membrane are caused by admixture of epinephrine contained in the original substance. An enzymatic decarboxylation in vitro could not be observed. On the other hand, the effects of the esters derived from the new epinephrine carboxylic acid were similar to those of the isopropylnorepinephrine. The most active ester derived from the new acid equalled isopropylnorepinephrine in its potency on all test subjects. In all experiments, the increasing effects on heart frequency, the bronchodilating and the depressing effect on the blood pressure of the esters run parallel. The corresponding derivatives of the p-oxyephedrin were practically completely ineffective. The connections between chemical constitution and pharmacological activity is discussed.

2695: H. G. Beljers: Ferrite isolators in the 8-9 mm wave band (Commun. Congrès int. Circuits et Antennes Hyperfréquences, Paris 21-26 Oct. 1957, part 2 (suppl. Onde électr. **38**, No. 376 ter), pp. 647-648, 1958).

For the construction of isolators at 8-9 mm wavelengths two main principles are applied. The first one is a Faraday rotator type with three ports, which is essentially a circulator. The reflected energy is dissipated in an external load in a waveguide normal to the main guide, decoupled from the input. The second type is a resonance isolator with a transverse magnetic field. The usual soft ferrite samples would require a magnetic field of about 12 000 Oe, which is rather high. If oriented anisotropic materials are applied the required field can be considerably reduced or even omitted. Very simple devices are resultant, although isolation ratios are generally less high than for the normal resonance isolators.

VOLUME 21, 1959/60, No. 9 pp. 253-280 Published 4th August 1960

Philips Technical Review

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

AN AUTOMATIC PARTICLE COUNTER AND SIZER

by H. A. DELL *), D. S. HOBBS *) and M. S. RICHARDS *).

531.717:531.791:621.317.39

Considerable attention has been paid in the last decade or so to the problem of automatically performing counting and sizing of samples of particles. This article gives an introduction to the problem and describes an advanced instrument for this purpose developed by the Mullard Research Laboratories.

In many investigations it is useful to be able to count and find the size distribution of numbers of particles [1]). The counting of blood cells, the measurement of the pigment granules in paints and the assessment of the explosion and health hazards of dust-laden coal-mine air are typical instances.

At some stage of each of these investigations a prepared sample from the medium in question is laid out on a flat surface. This two-dimensional specimen is then examined point by point by some microscopic or similar means. From the method of preparation and the features discovered in a known area of the two-dimensional specimen, the character of the original sample is determined.

It can be seen that the act of examining a plane field of view point by point is in many ways analogous to that of scanning a scene or picture as is done in television or phototelegraphic techniques. It is not surprising, therefore, that many methods based on these techniques should have been suggested for *automatically* analysing two-dimensional specimens of this type.

After a discussion of the nature of the problem, some of the methods used for the automatic counting and sizing of particles will be outlined. The methods used in the Mullard equipment will then be described more fully and further details given of the construction of this equipment. *Fig. 1* shows a photograph of the Mullard particle analyser.

Statement of the problem

For both the counting and the sizing of particles some type of scanning procedure is required. Furthermore it is necessary to have some method of ensuring that a particle is not counted more than once, or that two or more particles are not lumped together as one. When counting visually through a microscope, the observer's eyes perform the scanning function, while his memory of the counted particles in the whole field has to preclude duplicate or lumped counts. This can be very fatiguing and is not always very reliable, especially when the sample consists of a large number of particles (as is desirable to keep statistical errors low).

In an automatic apparatus for counting and sizing the scanning may be performed by a light beam, the modulation of which by the particles is detected by some type of photocell. Two systems that have been described are the line scan and the spot scan. In the former, a long narrow illuminated aperture is moved at right angles to its length, sweeping out an area as it does so. In the other, a small spot of light is moved over the specimen in a series of parallel lines until the required field has been covered. With the line scanning system duplication of counts gives less trouble than the counting of two or more particles as one, but certain artifices may be employed to

*) Mullard Research Laboratories, Salfords, Surrey, England.

[1]) For a comprehensive review of the whole subject of particle size analysis, see The physics of particle size analysis, Brit. J. appl. Phys., Suppl. No. 3, 1954. See also G. Herdan, Small particle statistics, Elsevier, Amsterdam 1953. A recent review of automatic sizing methods is given by B. B. Morgan, Research (London) **10**, 271-279, 1957.

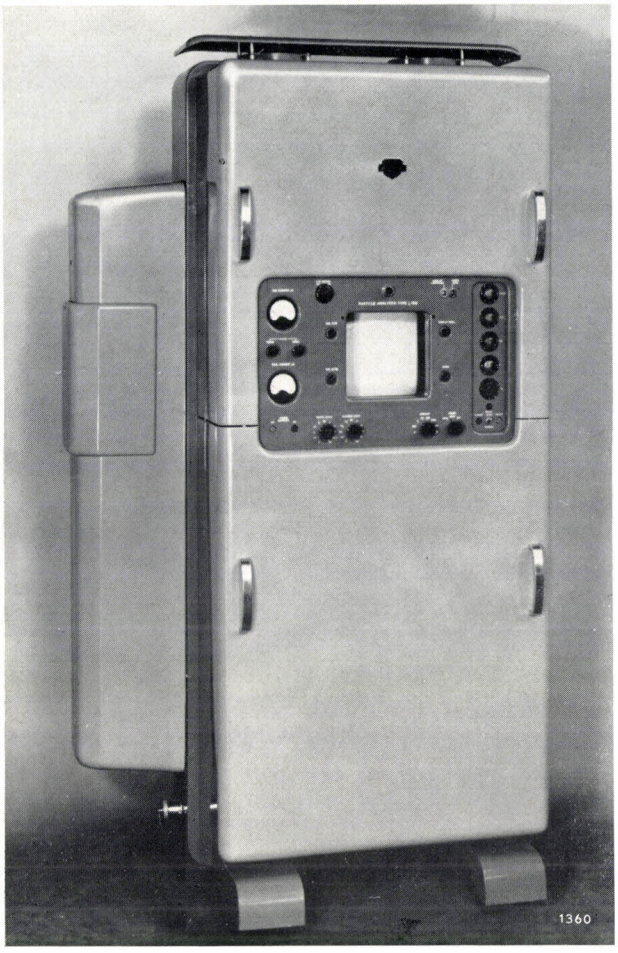


Fig. 1. Photograph of the Mullard particle analyser, type L 188.

reduce such errors. In the spot scanning system, duplicate, multiple or lumped counts can be largely eliminated with the aid of a line-to-line "memory".

For automatic *sizing* of particles each particle encountered has to be "explored" in some way and quantitatively assessed. In the Mullard instrument this is done by measuring the length of the intercepts of each particle. Another method is that whereby the scanning spot is momentarily deflected perpendicular to the scan direction as it encounters a particle and hence measures a "diameter" of the particle. Several other systems are possible. In all these methods, however, an important simplification is achieved by successive scanning of the whole field, each time for particles greater than a certain pre-set size. Each complete scan thus counts all particles greater than the pre-set size; from the successive scans an integrated size distribution curve can be plotted.

The accuracy of counting and sizing depends on the resolution of the instrument in relation to the particle size, the size of the field in relation to the particle size (edge effects) and the number of particles in the field (sampling errors and edge effects). Clearly these various errors can be diminished by examining different parts of the sample and by multiple sampling.

Scanning of the sample

The line scanning method is illustrated diagrammatically in *fig. 2*. An idealized specimen containing three particles A, B and C is shown at (a), while (b) indicates the light signals that would be generated by such a specimen as it was scanned.

It can be seen that while particle A generates a unique signal from which its dimensions could be determined, particles B and C generate a combined signal, from which their individual dimensions, or even their separate presence, cannot be deduced directly. If all individual particles are of interest, it is necessary then to make sure that only one is intercepted by the illuminated aperture at once, or to include some statistical correction for multiple interceptions.

In the same way, particles only partially intercepted at the ends of the line cause inaccurate signals, as they are indistinguishable from smaller particles fully intercepted.

Two methods have been suggested for dealing with this latter effect. One is to perform two (or more) experiments using different lengths of swept line [2]). If it can be assumed that the end effects are statistically the same in both instances, then the difference between the results will give the effect of sweeping the specimen with a line of length equal to the difference between the two lengths actually used but possessing no end effects. When such a system is used it is important that the larger line should not substantially alter the chances of intercepting more than one particle at a time, as may easily occur with non-uniform specimens.

The other method suggested and also used in a commercial instrument [3]) uses a triple swept line system. In this, the main measuring line has associated with it two short guard lines, one at either end. They and the measuring line form a single exploring system which is swept across the specimen. The circuit logic ensures that particles intercepted by the main line and one, or both, of the guard lines are disregarded. Thus only particles intercepting the main line alone are recorded by the apparatus. This technique suffers from the disadvantage that the exact area of the specimen effectively scanned is larger for small particles than for large ones, so that once again a measure of uncertainty is introduced.

In view of these shortcomings the Mullard particle analyser uses the spot scanning method (*fig. 3*).

[2]) H. S. Wolff, Nature **165**, 967, 1950. A commercial instrument using this system is described by P. G. W. Hawksley *et al.*, Brit. J. appl. Phys., Suppl. No. 3, p. S.165, 1954, and B. B. Morgan and E. W. Meyer, J. sci. Instr. **36**, 492, 1959. The theory of the method was given by P. G. W. Hawksley, Brit. J. appl. Phys., Suppl. No. 3, p. S.125, 1954.

[3]) E. W. Meyer, British Patent 714 350, 1951. Edge effects are also discussed by H. Nassenheim, Chem. Ing. Tech. **27**, 38-39, 535-542, 787-794, 1955.

N.V. PHILIPS' GLOEILAMPENFABRIEKEN

A specimen containing four particles *A*, *B*, *C* and *D* is shown at (*a*), while (*b*) indicates the signals that are generated when the specimen is scanned by a very small spot.

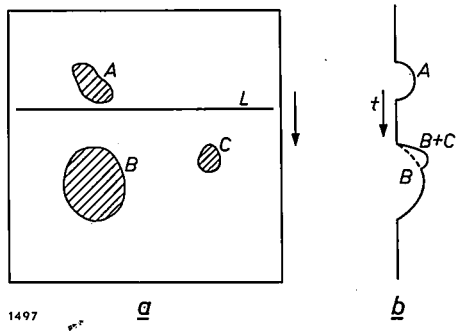


Fig. 2. Illustrating the *line* scanning method. *a*) Idealized particle sample. *L* is a line of light moving in the direction of the arrows. *b*) Signals obtained as the scanning proceeds.

In this case the source of any signal can always be identified uniquely with a definite location in the image plane, determined by the spot position. Within certain limits of resolution determined by the finite spot size and scan line spacing, neighbouring particles can always be separately distinguished. In fact, within these limits, the signals generated must contain all information about the particles present, for a specimen image can be reconstructed from them (cf. facsimile telegraphy and television). Unfortunately, information about particle numbers and sizes can be extracted only with some difficulty.

The reason for this is that large particles overlap several lines of scan and so produce signals on each interception. If the size of each particle were required, it would be necessary in addition to store up all the intercept information due to any one particle. On completing the scan either the number or the magnitude of the intercepts on each particle could then be determined. (As will be seen presently, the present instrument sizes by intercept "maximization"; this does not require storage of *all* the intercept information from each particle but only line-to-line storage.)

If the nature of the specimen is well established beforehand (in particular if the shape of the particles is more or less invariant, e.g. circular), this recognition of multiple interception can be based purely on a statistical analysis related to intercept length [4]). It is then only necessary to count and measure all the interceptions found in a field of view to be able to infer the number and size distribution of the particles present. More often, as the very nature of the specimen is the point in question, this is impossible. A more precise recognition of signal association then becomes necessary.

Line-to-line signal association

In order to count and size in an automatic machine, it is necessary to lay down some *criterion* by which successive interceptions may be judged as belonging to the same or different particles.

Two different criteria have been considered in the literature. The first, called the "anticipation-tolerance" criterion [5]), states that a signal in any line of scan must be regarded as due to a large particle overlapping the previous line of scan only if a signal occurred in the previous line whose leading edge (say), as measured along the line, lay within a predetermined tolerance distance from the leading edge of the later signal.

In the instance illustrated (fig. 3), a tolerance may be selected which would, for example, correctly associate the first signal in line *8* with that in line *7*, but which would rightly exclude the second signal in line *8*, recognizing it as due to a separate particle. If particles exist having extended edges nearly parallel to the lines of scan, a large tolerance is necessary. It may then be very difficult to avoid incorrectly associating signals due to separate but nearby particles. Similar difficulties arise if trailing-edge signals are used.

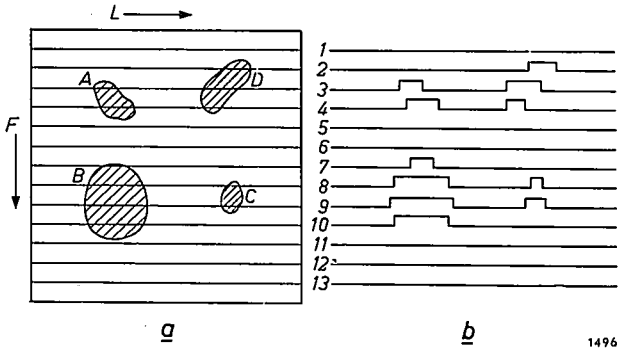


Fig. 3. Illustrating the *spot* scanning method. *a*) Idealized sample. *L* is the line scan direction and *F* the frame scan direction. *b*) Signals obtained from the particles shown.

The second, the "overlap" criterion (the advantages of which were first emphasized by Roberts [6])), states that a signal in any line of scan must be regarded as due to a large particle overlapping the previous line of scan only if a signal occurred in the previous line *some part of which* lay the same distance

[4]) J. G. Dawes and P. G. W. Hawksley, reported in W. H. Walton, Central Research Establishment Report of National Coal Board (U.K.) No. 79, 12 June 1951.

[5]) D. W. Gillings, British Patent 719 773, 1950.

[6]) F. Roberts, reported in W. H. Walton, Nature **169**, 518, 1952.

along the line as *some part* of the later signal (i.e. some signal overlap existed). When a light spot of finite size is used to scan lines spaced by approximately the spot *radius*, it can be shown that the overlap criterion will correctly interpret all signals down to a resolving limit of about the spot *diameter*. The false association possibility inherent in the anticipation-tolerance criterion, which may become very significant when dense specimens are considered, is small in this case.

The overlap criterion is the superiorw hen counting only is required, but when *sizing* by intercept *summation* it runs into difficulties. For this reason, all the intercept-summing sizing systems so far described in the literature use the anticipation-tolerance criterion [5] [7] [8].

Consider, for example, the signals due to the particle *D* in fig. 3. In the second line of scan of the field the first intercept signal from this particle occurs and must be suitably stored away. In the third line of scan, the second intercept signal will occur, but only after a certain delay (here just over half the duration of this signal) will association with the first intercept be recognized. Until that moment it could have been due to a new and different particle. All of the signal must be stored correctly, however, for otherwise a later signal, such as that in the fourth line, may never be recognized as due to part of the same particle. Some temporary store is thus necessary in which all signals can be held until their ultimate destination is known.

This difficulty does not arise with the anticipation-tolerance criterion. A signal in any line of scan either lies within the tolerance zone of a previous interception, when the correct destination is known, or it lies outside all such zones, when it must represent a new particle. In either case the appropriate action is established from the *beginning* of the signal.

The present instrument, however, sizes by intercept "maximization", that is, each intercept pulse is compared in length to a certain pre-set value and a count is made only when this pre-set value is exceeded [9]. As a result it is only necessary to store intercept information for the duration of *one* line of scan, and there is then no fundamental difficulty in applying the overlap criterion.

Spot scanning systems in general suffer from the disadvantage that an indented particle may be met in such a fashion that it initially appears to the scanner as two or more separate particles (see *fig. 4*). Only when the full bulk is discovered may their unity be recognized. This difficulty is not easily overcome, although the errors it causes are not so serious because such indented "particles" are often, in fact, agglomerates.

Line-to-line memory

Whichever criterion of signal association is used, some form of memory is necessary in which it is possible to store the signals generated by the scanning system. During the scanning of any line this record must then be read, so that a representation of the signal generated one line previously is available.

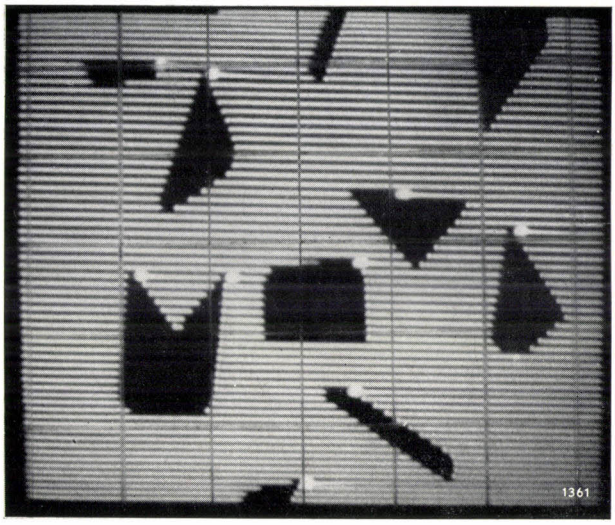


Fig. 4. Scanning of idealized particles, displayed on a monitor tube. Direction of scanning; right to left. The white dots are marker tags to indicate when a count has been made (see below). It will be noted that the indented particle (lower left) has been counted twice.

In a particle analyser these remembered signals are compared with the output of the scanner. For every particle, only on the first intercept does a scanner signal exist which has no counterpart in the memory. This condition may be made to generate a unique impulse corresponding to each particle. (During scanning of the line following the last intercept there also occurs a signal unique to each particle; in this case it is the delayed signal from the previous line, the direct signal being zero.)

In addition to true line-to-line memory systems a double-spot pseudo-memory [10] may be used. In this the specimen is scanned by two similar but identifiable light spots and the light signals transmitted at the two scanning points are indicated in two separate channels. If as one spot scans the n-th line of the raster and the other (trailing) spot similarly scans the $(n-1)$th line, signals are available in the two channels which correspond to those which a single-spot scanner and a memory would produce.

This simple method of imitating the action of a memory suffers from a practical disadvantage, first pointed out by

[7] C. F. Bareford and H. A. Dell, British Patent 727 134, 1951.
[8] W. H. Walton, D. G. A. Thomas and J. W. Philips, British Patent 732 662, 1951.
[9] D. S. Hobbs, British Patent Appl. 16101/56.
[10] H. A. Dell and E. Jones, British Patent 741 471, 1951.

Walton (private communication), which is more serious than would appear. The signal delivered by the trailing spot is not the remembered signal which the leading spot delivered; it is instead a "repetition" of that signal produced by the trailing spot. If, therefore, the trailing spot either takes a slightly different path from the leading spot, or possesses a slightly different signal sensitivity (both conditions almost impossible to avoid in practice), the reconstructed signal may be false. The result of this potential fallibility is that in many instances particles may be counted twice or missed altogether.

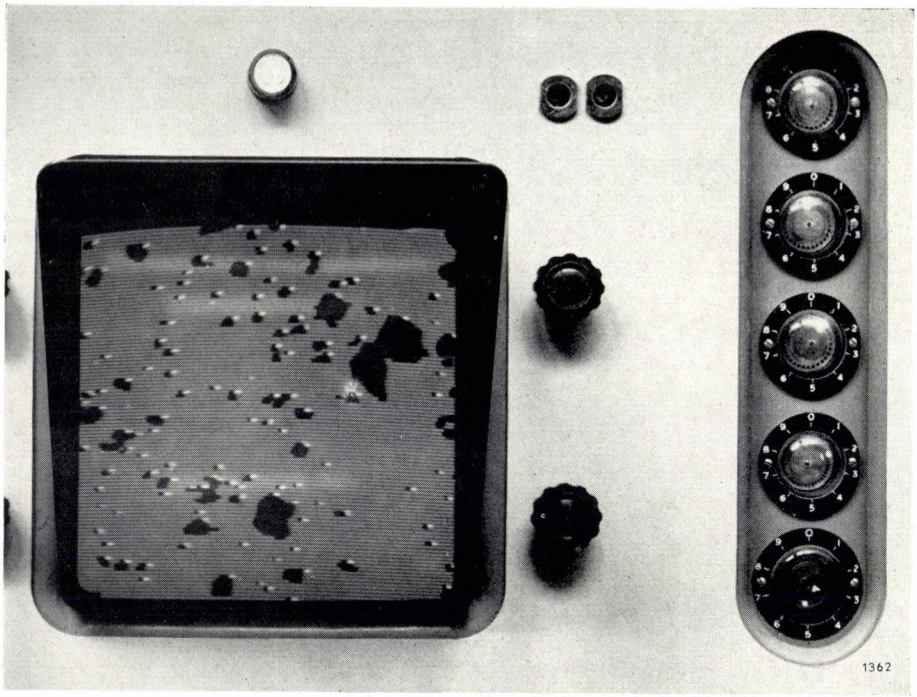


Fig. 6. Display panel showing the 12″ monitor tube and the dekatron counters. As each particle is counted it is "tagged" with a bright spot by a strobe pulse from the counter. The photograph was taken with an exposure time corresponding to more than 20 frames. The sharpness of outline indicates the stability of the instrument.

Many different forms of line-to-line signal memories have been tried. These include magnetic drum and tape recorders, electrostatic cathode-ray tube storage systems [11]), acoustic delay lines of the wire type [12]) and mercury delay lines. Of all these, mercury acoustic delay lines have proved the most satisfactory for this application. Such a delay line, providing a delay of 1000 μsec, is employed in the Mullard instrument for the line-to-line memory and will be described later in this article.

The Mullard particle analyser

In the Mullard instrument a photograph of the sample is first made on standard 35 mm film and it is this photograph that is examined by the instrument. The philosophy behind this is that it leaves the user greater freedom in sampling methods and in exploiting the various techniques available to the microscopist. This makes the instrument more versatile, for it would be impracticable to build in elaborate optical facilities into what is already quite a complicated instrument, quite apart from the costs involved and the undesirable duplication of equipment already available in most laboratories.

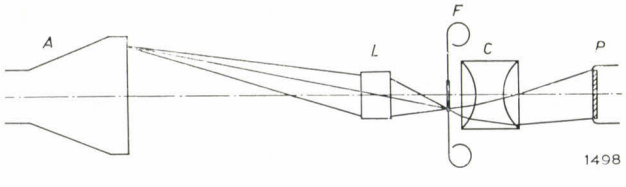


Fig. 5. Flying-spot scanner showing the optical system which reduces the 60 × 60 mm raster on the cathode-ray tube to a 20×20 mm image on the film. *A* scanner tube. *L* projection lens. *F* film sample. *C* condenser. *P* photomultiplier. In the actual instrument two projection lenses are provided in a turret mount, the second lens giving a reduction of about 12×. The raster image on the film is then only 5 × 5 mm, so that the resolution is higher.

The film is scanned by a flying-spot scanner [13]). A scanning raster of 100 lines, 60 × 60 mm is produced on the face of a cathode-ray tube, type MC 13-16, which has a very short-persistence phosphor [14]). The frame frequency is nearly 10 c/s, so that the whole field is scanned in about 1/10 sec. The raster is projected in reduced format by an optical system (*fig. 5*) onto the 20 × 20 mm image on the film. The transmitted light falls on a photomultiplier tube producing a signal related to the local opacity of the area scanned at any moment.

For monitoring purposes, a 12″ cathode-ray tube displays the scanned area at a linear magnification of 10× (*fig. 6*). Each particle is "tagged" with a bright marker spot as it is counted. When

[11]) F. C. Williams and T. Kilburn, Proc. Instn. Electr. Engrs. **96-III**, 81, 1949.

[12]) E. M. Bradburd, Electrical Communication **28**, 46, 1951.

[13]) The principles of this type of scanner are outlined in F. H. J. van der Poel and J. J. P. Valeton, Philips tech. Rev. **15**, 221, 1953/54.

[14]) A. Bril, J. de Gier and H. A. Klasens, Philips tech. Rev. **15**, 233, 1953/54.

sizing, the bright tags appear only on those particles with intercepts greater than the pre-selected value. With the normal projection lens (20×20 mm raster on film), the size pre-set control gives a series of eleven critical sizes, increasing by a factor $\sqrt{2}$, viz. 0.2 mm, 0.28, 0.40, 0.56, 0.80, 1.14, 1.60, 2.28, 3.20, 4.56, 6.40 mm.

The results of counting or sizing appear on decade counters situated on the display panel (fig. 6). The count appearing there is in fact the totalled count obtained by scanning the field ten times, the repetitive scanning being performed automatically. The reading therefore has to be divided by ten. The reading is displayed for seven seconds, after which the cycle of ten scans repeats automatically. Alternatively, the reading can remain in the counters until reset manually after changing the sample or changing the pre-set size.

With the aid of the simplified block diagram in *fig. 7*, we shall now describe the operation of the particle analyser, leaving certain circuits and components to be described later when their functions and critical requirements will be more evident.

In order to understand how the instrument works, we must first consider in detail the system used for line-to-line signal association. It will then be clear how the instrument makes a simple *count* of the particles of a sample. To understand how the instrument measures a *size distribution*, we must further consider details of the sizing discriminator and certain associated circuitry which ensure that counts are registered only for particles greater than a certain pre-set size.

To simplify the explanations, we consider the pulses from only a few representative particles over a small portion of the scan.

Counting

Referring to the block diagram of fig. 7, we imagine the two-pole switch S to be in the right-hand position ("Count"). Of the four channels carrying the shaped intercept signals, only *1* and *2* are now

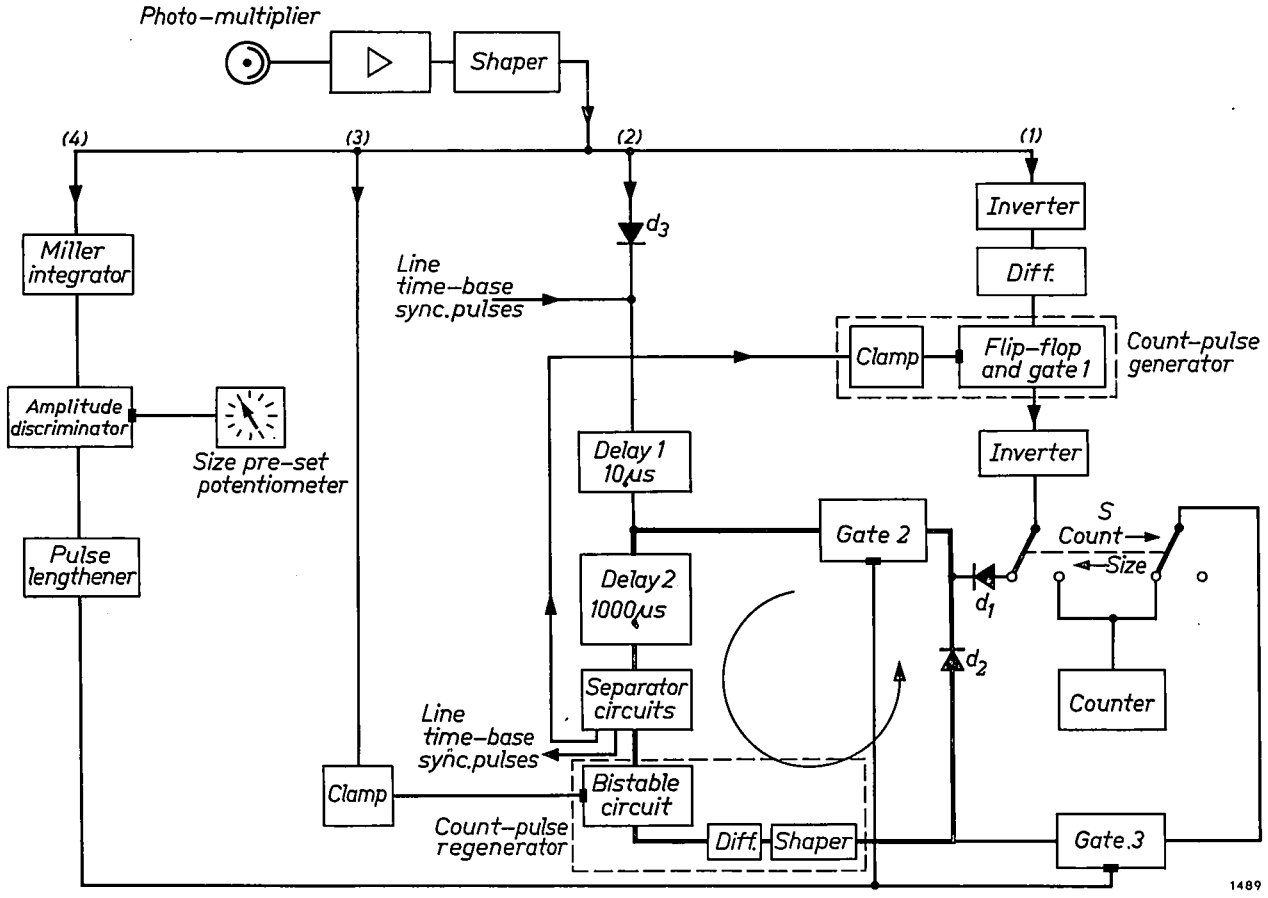


Fig. 7. Block diagram of the particle analyser. The particle-intercept signals produced in the photomultiplier of the flying-spot scanner are amplified, and then shaped by clipping. The shaped pulses are then fed to four channels: (*1*) the count-pulse channel, (*2*) the delay-line channel, (*3*) the recirculation-clamp channel and (*4*) the size-discriminator channel. During *counting* (switch S to right), only channels *1* and *2* are performing any function. In *sizing* (S as drawn) all four channels are functioning. Delay *1* is a lumped-parameter delay line; delay *2* is a mercury delay line.

operative. In channel *1* each intercept signal is inverted and differentiated. The resulting pulses are fed to gate *1* (which is always closed to negative pulses). The positive (trailing edge) pulse forms the "count pulse" and is allowed to pass gate *1* to the counter only if this gate has been opened [15]) by the leading-edge pulse operating an Eccles-Jordan flip-flop (see *fig. 8a*); the latter, however, can be operated only provided that it has not been clamped by an in fact cuts off its tail so as to give the count pulse a duration of about 3 μsec. On the second and succeeding intercepts of a particle (fig. 8*c*) the flip-flop is clamped by the presence of a delayed intercept pulse carried over from the previous line of scan, some part of which overlaps some part of the direct intercept pulse: *either* the delayed signal overlaps the leading edge of the direct signal *or* the direct signal overlaps the leading edge of the delayed signal [16]).

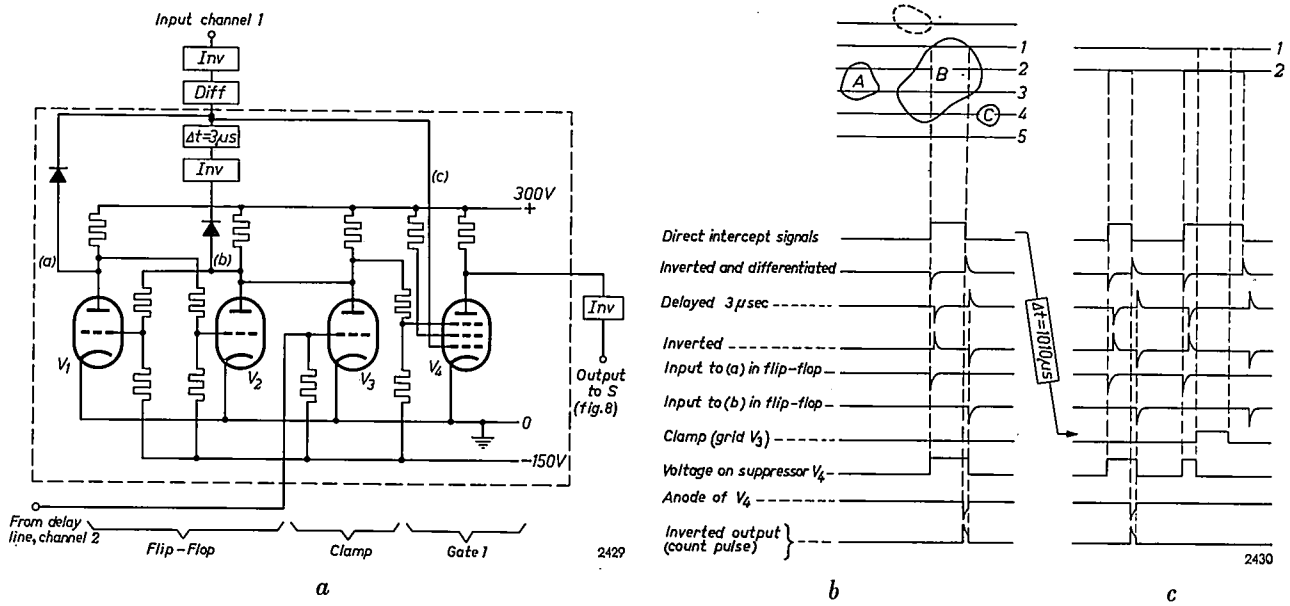


Fig. 8. *a*) Count-pulse generator (channel *1* in fig. 7) consisting of flip-flop, clamp and gate. The incoming shaped intercept signal is inverted and differentiated. The two peaks corresponding to leading and trailing edges are then fed to the three inputs *a*, *b* and *c* of the circuit. In the quiescent state (no intercept signal), V_1 is cut off and V_2 is conducting, so that V_1 is cut off.
b) Waveforms at various points during scanning of the line marked *1*, when the *first* intercept of a particle *B* occurs.
c) Waveforms during line *2*, when the first intercept of *A* occurs and the *second* intercept of *B*. Owing to the presence of the overlapping clamp pulse, delayed from line *1*, the line *2* count pulse of *B* cannot pass gate *1*. Likewise, on the succeeding lines, neither *A* nor *B* can be counted again. In line *4* the particle *C* will be counted since it is not overlapped by *B* in the *previous* line of scan. The dotted particle is drawn in the critical position of zero overlap with respect to *B* (see text).

intercept signal from the previous line via channel *2*, delayed by 1010 μsec (the total line scan period). It is here that the "overlap" criterion (see p. 255) is introduced. Normally, during the *first* intercept of any particle, there is no delayed pulse emerging from the 1010 μsec delay line. Hence the flip-flop remains unclamped and is triggered by the negative-going pulse corresponding to the leading edge of the intercept (fig. 8*b*). In this state the flip-flop applies a potential to gate *1* which opens it to positive pulses. The flip-flop is reset by the delayed (3 μsec) and inverted pulse corresponding to the trailing edge of the intercept. The 3 μsec delay makes sure that the count pulse has time to pass through the gate, and In the former case the negative (leading edge) pulse is unable to trigger the flip-flop and gate *1* remains closed. In the latter case the flip-flop is operated and the gate opened but before the count pulse (trailing edge of intercept) arrives, the clamping pulse has already returned the flip-flop to its initial state and therefore closed the gate. In neither case, therefore, does the count pulse pass gate *1*. Only on the first intercept of each particle does the count pulse pass through to the counter, giving only one count per particle, as required.

Now suppose there is an adjacent particle (dotted in fig. 8*b*) intercepted by the line previous to the first line of scan (*1*) of a given particle (*B*). This clearly

[15]) Used throughout this article in the sense that pulses can pass through a gate; when a gate is shut, the pulses are blocked.

[16]) Except in the case of a very slender particle making an angle of less than 45° with the scan direction.

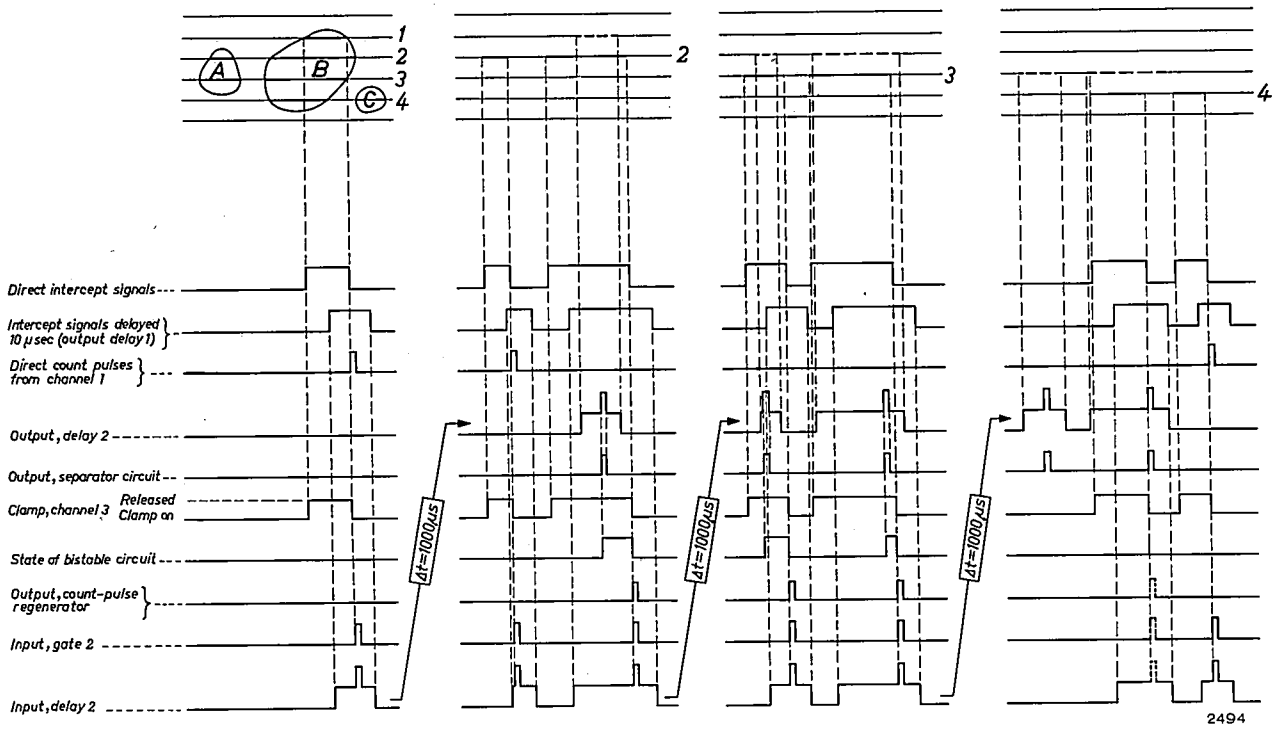


Fig. 9. Waveforms illustrating the regeneration and recirculation of count pulses. Portions of four lines of scan are shown, side by side, covering the three particles *A*, *B*, *C*. Note that delay *2* delays the pulses by 1000 μsec, that is, 10 μsec *less* than the line-scan period.

Between lines *3* and *4* the shape of the exit-edge of particle *B* is such that the clamping pulse in channel *3* barely overlaps the recirculating count pulse of this particle. The bistable circuit in channel *2* is therefore only just — or just not — triggered, so that the count-pulse regenerator may not always emit a regenerated count pulse (dotted in column *4*). This gives rise to negligible sizing errors, however, since the maximum intercept of most particles occurs before such an exit-edge is encountered; any particle greater than the pre-set size has then already been counted and its count pulse has ceased to recirculate.

gives a delayed clamping pulse close in time to the first direct intercept pulse from *B*. Provided that there is no actual overlap of these two pulses, the above circuitry correctly interprets *B* as a new particle and makes a count. If, however, the pulses overlap, *B* would be interpreted as an extension of the particle shown dotted. Only one count would thus be made for the two particles.

In counting, the mercury delay line carries only two sets of signals, viz. the intercept signals from channel *2* (fig. 7) and the time-base synchronizing pulses (see later). The separator circuits following this delay line (see below) simply route the delayed intercept signals to the count-generator clamp and separate out the synchronizing pulses for the scanner time-base; only during sizing is there a third set of pulses ("count" pulses in the recirculation circuit, bold lines in fig. 7) to be routed by the separator circuits. We now consider the sizing operation.

Sizing

For measuring the size distribution of a particle sample, the switch *S* is placed in the left-hand position ("Size") as drawn in fig. 7. All four channels of fig. 7 and also the recirculation circuit (bold lines) are now operative. The trailing edges of all intercept signals generate "count" pulses in channel *1* but, again, only the *first* count pulse of each particle passes gate *1*. With *S* as drawn, each such pulse (length 3 μsec) now passes via a diode to gate *2*, which is normally open. It then enters the 1000 μsec delay line, being superimposed upon the shaped intercept signal, itself delayed by 10 μsec (*fig. 9*). After separation from the intercept and sync pulses (see later), the count pulse passes to a bistable circuit (flip-flop) (fig. 7), which it triggers provided that the clamp has been released by the leading edge of a direct intercept signal to the clamp via channel *3*. (A direct intercept signal is always present except during the scanning line *following* the last intercept of a particle; what happens then we shall see presently.) The bistable circuit remains in its triggered state until reset by the re-application of the clamp at the end of the intercept signal.

When the bistable circuit resets, a differentiating circuit produces a pulse which is shaped to form a regenerated count pulse. Gate *3* being normally closed, this count pulse again passes (now via diode

d_2) to the gate *2* (normally open) and hence recirculates round the path shown in bold lines in fig. 7. Trains of count pulses, having originated at the first interception of each particle and being regenerated once per circulation, are therefore circulating continuously round this path while the corresponding particles are being scanned. Each count pulse continues to recirculate until the length of an intercept pulse of the associated particle exceeds a certain pre-set value, when gate *2* is momentarily shut and gate *3* simultaneously opened; this particular count pulse then passes to the counter via *S* and its recirculation ceases. This action is performed by the sizing discriminator (channel *4*), to be described presently.

If, however, the first or succeeding intercepts of a given particle never reach the pre-set pulse length, recirculation of the count pulse is stopped when the scan of that particle is complete, and it is not allowed to proceed to the counter. This is achieved by the clamp in channel *3*, mentioned above. The count pulse of the undersize particle, initiated at the first intercept, is recirculated (and regenerated) during each line of scan in the manner described above. After the last scan of the particle, however, there is no direct intercept signal in channel *3* to release the clamp on the bistable circuit at the moment that the delayed count pulse arrives. For this reason the bistable circuit cannot be triggered by the incoming count pulse. Since it is not triggered, there is also no reset and hence no regenerated count pulse is passed on. Recirculation therefore ceases and the undersize particle is not counted.

We must now examine how the instrument discriminates between particles greater than the pre-set size and those smaller. The pre-set signal length is chosen by setting a potentiometer in the discriminator in channel *4* (fig. 7). The discriminator (*fig. 10a*) consists of two parts: (i) a Miller linear ramp generator driven by the direct intercept signals, and (ii) a Schmitt trigger circuit acting as a pulse amplitude discriminator. The former is a Miller integrator circuit in which the charge and, therefore, the voltage on a condenser is made to depend linearly on the time for which the intercept signal is present at its input. A pulse *amplitude* discriminator can thus be used to discriminate between intercept signals

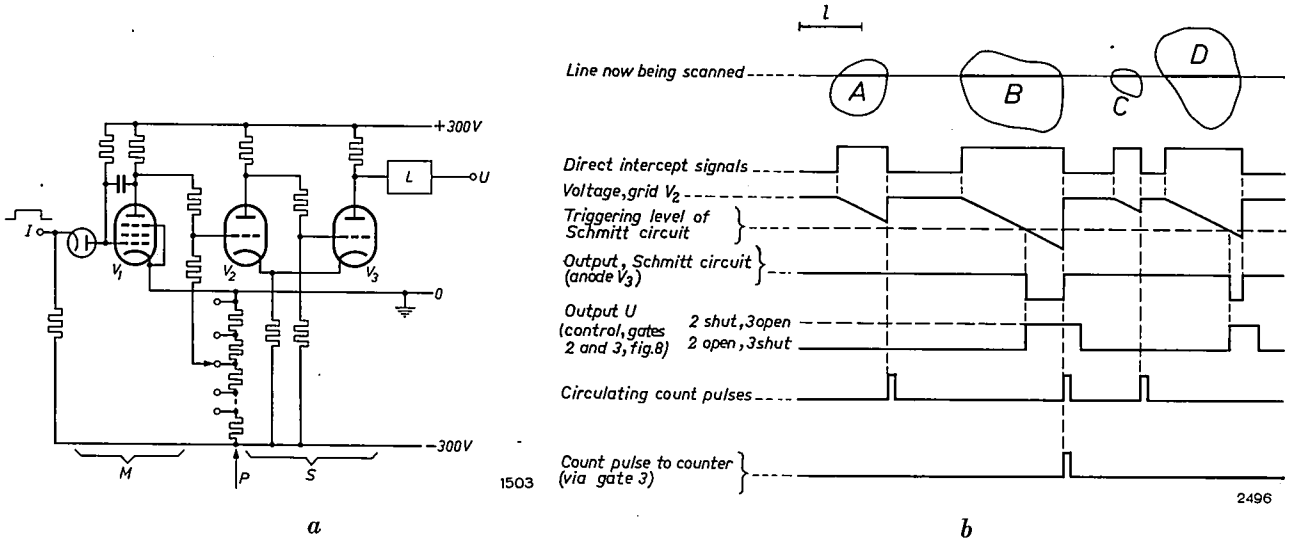


Fig. 10. *a*) Sizing discriminator circuit (simplified). Direct intercept signals are fed via channel *4* (fig. 7) to the input *I* of the Miller integrator *M*, and cut off the diode for the duration of each signal. The anode voltage of V_1 is linearly dependent on the duration of each signal. At a certain value of the grid voltage of V_2, set by the size pre-set potentiometer *P*, the Schmitt circuit *S* is triggered. The pulses at the output *V* are lengthened by 8 μsec by a monostable flip-flop *L*.
b) Waveforms during the scanning of part of a line. Four typical particles are shown. With the potentiometer set for the particle size *l* indicated, *A* is too small to be counted but will be counted in the next line of scan; the *B* intercept is greater than the pre-set size, so that *B* is counted; particle *C* is too small and will not be counted; the *D* intercept is greater than the pre-set size, but this particle has been already counted in a previous line, so that its count pulse has ceased to recirculate.

longer or shorter than a given pre-set value. This value is set by means of a potentiometer which controls the grid bias of the input valve of the Schmitt trigger circuit. As will be clear from the waveforms in fig. 10*b*, the latter emits a square pulse whenever the Miller voltage exceeds the pre-set value necessary to trigger it. This is the case whenever a direct intercept signal is longer than the pre-set value. The duration of the pulse from the

Schmitt trigger is governed by the time for which the Miller voltage remains above the pre-set value, i.e. until the end of the direct intercept signal; however, the Schmitt circuit is followed by a monostable flip-flop which adds a further 8 μsec to the length of this pulse. In this way the pulse emitted from the discriminator overlaps by some 8 μsec the trailing edge of the corresponding intercept signal. The discriminator output pulse passes to gate *2* which it shuts and gate *3* which it opens. Hence the recirculating count pulse of this particle (which has just been regenerated by the reset action of the signal via channel *3*) ceases to recirculate, but passes via gate *3* to the counter.

It will be clear that for the correct functioning of the circuits it is essential that the total line-scan period (scan + flyback periods) always remains exactly equal to the total delay provided by delay lines *1* and *2*, viz. 1010 μsec (nominal); if it were not, the delayed signals would not always be correctly phased with respect to the next line of scan. It is therefore necessary to lock the line time-base to the actual delay period. This is done by passing the time-base triggering pulses (sync pulses) through the delay lines, as outlined presently.

Further details of various components and circuits

Having described the essentials of the equipment and how it counts and sizes, we now go back to consider the design and operation of one or two of its critical units, in particular the delay line, the separator circuits and the time-bases and their synchronization.

The complete instrument with covers removed is pictured in fig. 11; the location of some of the sub-units is given in the caption.

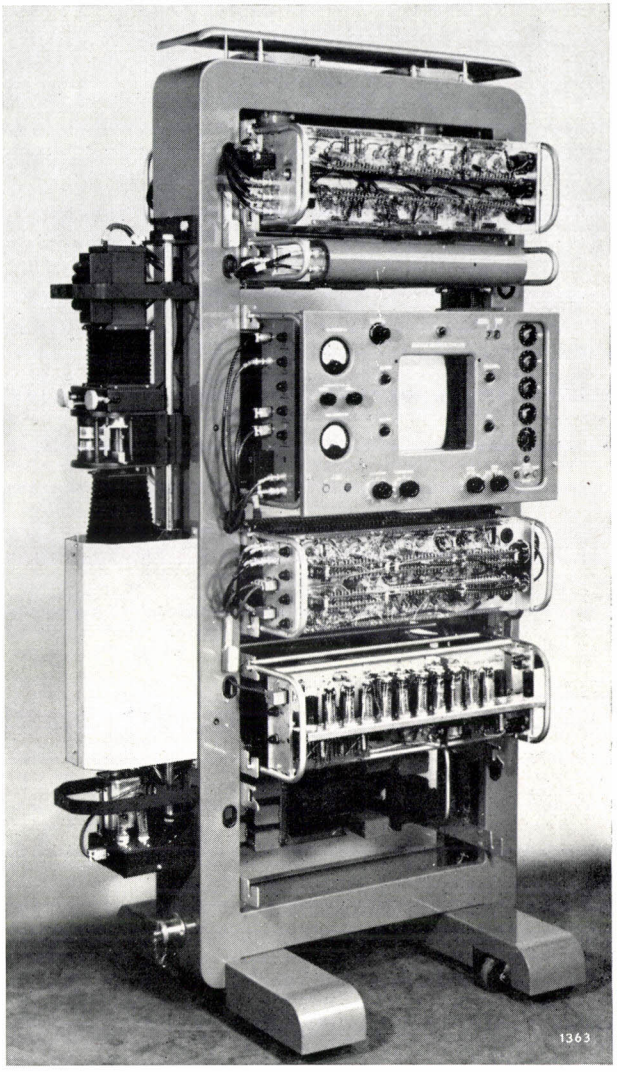


Fig. 11. Complete instrument with covers removed. On the left is the flying-spot scanner. The scanning tube is concealed by a screen (left). Between the two bellows can be seen the two alternative projection lenses (cf. fig. 5). The units visible are, from top to bottom: video chassis, containing intercept-pulse shaper and amplifier, count-pulse generator and delayed intercept-pulse separator circuit; mercury delay line and associated circuits, including sync-pulse separator; display panel; line and frame time-bases and synchronizing circuits; power pack.

Scanning raster, resolution and bandwidth requirements

In planning the instrument it was decided that the smallest particle to be measured should have a diameter of 1% of the line-scan length. This gives a sufficient resolution for most applications. For the resolution in the vertical sense, assuming a square field, a minimum of 100 lines is indicated.

The frequency of frame scanning should not be too low, otherwise display difficulties will be encountered. On the other hand it should not be too high, otherwise expensive high-speed logical circuits will be required to be able to cope with the above-mentioned resolution. With a 100-line raster, a frame frequency of about 10 c/s was chosen; the flicker of the display is then not excessive and focus and clipping adjustments can be made reasonably rapidly [17]). The line-scan period is 1010 μsec, as mentioned earlier; allowing a 30% blanking period for flyback and edge blanking, the active trace period is 700 μsec, making the smallest element to be resolved 7 μsec (1% of picture width). The signal channels (photomultiplier, video amplifier, switching circuits, delay line) must therefore be able to handle pulses of 7 μsec without introducing excessive distortion.

The requirements of the flying spot on the scanner tube are: *a*) small enough spot to give the required

[17]) A later version of the instrument employs a 120-line raster and a frame frequency of ~7 c/s.

resolution, and *b*) sufficient light output. The effective diameter of the spot depends on the tube beam current and the photomultiplier clipping level (*fig. 12a*). With the MC 13-16 tube and the gain employed the effective spot diameter d is about 0.3 mm on the screen. The raster being 60 × 60 mm, the spot represents about 0.5% of the active line scan and is therefore small enough to resolve the smallest particles (1% of line scan). Assuming that the intensity of the spot is constant over its effective diameter d,

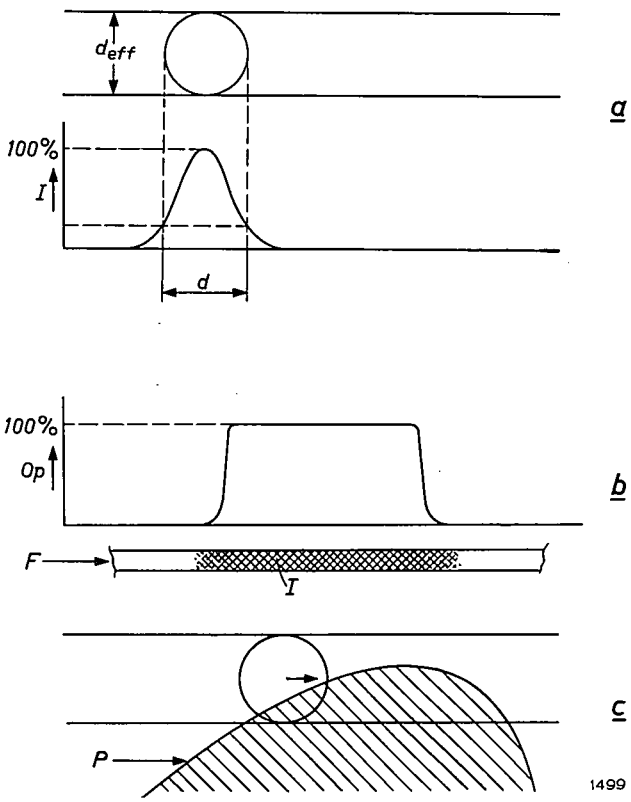


Fig. 12. *a*) The effective diameter d_{eff} of the scanning spot varies with the scanner-tube beam current and the photomultiplier gain, owing to the non-uniform intensity profile of the spot.
b) Percentage opacity *Op* of a particle image *I* in the photographic film *F*.
c) Oblique encounter of scanning spot with a particle edge *P*.

the rise time τ of the video intercept signal when the spot crosses a particle edge (assumed perfectly sharp and perpendicular to the direction of scan) is

$$\tau = td/l,$$

where t is the period of the active line scan and l its length. With $t = 700$ μsec and $l = 60$ mm, we find $\tau = 3^1/_2$ μsec. The video amplifier system was therefore to be designed so that pulses with rise times of rather less than this (2 μsec) pass through with little deterioration while not allowing excessive noise to pass. The theoretical bandwidth required [18]) is

$$\beta \approx 0.4/\tau = 0.4/(2\times 100^{-6}) = 200 \text{ kc/s}.$$

Final trimming of the amplifier stages was done empirically, using a test sample photographically reproduced from drawings.

The rise-time of the particle intercept signals is in general somewhat longer than the figure given above for the following reasons:

a) the intensity of the spot is *not* constant over its effective diameter as assumed above; moreover the effective diameter varies somewhat with fluctuations in the photomultiplier and amplifier system (fig. 12*a*);
b) the edge of the photographic image of each particle has a certain opacity profile (fig. 12*b*);
c) in general the particle edge is met obliquely by the scanning spot (fig. 12*c*).

As regards the spot brightness, the permissible energy in the raster is related to the area of the tube face actually scanned. With 100 lines, each of about 0.3 mm width, the area covered is effectively only about 50% of the 60 × 50 mm raster. The tube is therefore run at 20 to 50 μA (max.) beam current, i.e. about 50% of its rated value. This gives sufficient light, while helping to ensure a small spot diameter.

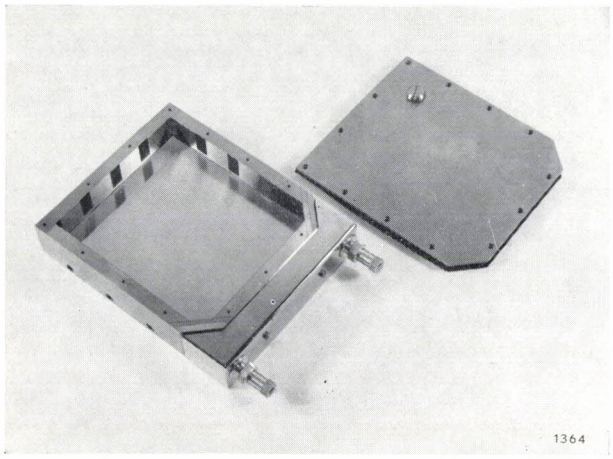
Mercury delay line

The mercury acoustic delay line [19]) consists of a sealed rectangular tank of mild steel (*fig. 13a*) about 2 cm deep, filled with mercury (filled under vacuum to exclude all air and water-vapour). Identical quartz crystal plates (X-cut) mounted at the 45°-bevelled corners, function as transmitting and receiving elements. Each plate, vibrating in the thickness-compression mode and operating at a wavelength very small compared to its diameter, has a polar diagram with a very narrow main lobe (first minimum at ~ 0.7°). With such a narrow ultrasonic beam and such an angularly-selective receiver it is possible to allow the beam to make multiple reflections at the walls before arriving at the receiver crystal. This makes for a compact delay line. Fig. 13*b* shows the path of the main beam.

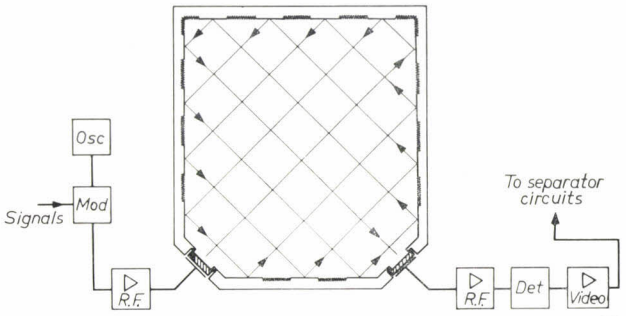
To reduce the small amounts of coherent sound energy reaching the receiver via other paths (e.g. signals emitted by the side lobes), the regions of the walls between the reflection points of the main beam are grooved, so that local specular reflection is prevented. Unwanted ultrasonic radiation is therefore scattered, destroying any phase coherence. Spurious signals originating mainly in the side lobes are thereby largely eliminated.

---

[18]) See, for example, G. E. Valley and H. Wallman, Vacuum tube amplifiers, p. 80 (Rule 4), M. I. T. Radiation Laboratory Series, Vol. 18, McGraw-Hill, New-York 1948.
[19]) C. F. Brockelsby, Ultrasonic mercury delay lines, Electronic and Radio Engr. **35**, 446-452, 1958.



a



b

Fig. 13. *a*) Photo of delay-line tank. Internal dimensions 115 × 102 mm.
b) Sketch of delay line with associated units. The transmitted beam undergoes 15 reflections before arriving at the receiver crystal. Between the points of reflection, the walls of the tank are grooved (angle of groove ≠ 90°) to scatter any off-beam coherent radiation.

Another source of spurious signals is that portion of the sound energy of the main beam which is reflected by the receiver crystal, returns to the transmitter and is reflected back to the receiver again. These "third-time-round" signals (with 3× the normal delay) can be reduced to negligible amplitudes by tilting the receiver crystal through an angle Θ (rather more than 0.2°) from its optimum position. This causes only a slight reduction in the sensitivity to the main beam, but the third-time-round sound is incident at an angle 3Θ ($\approx 0.7°$), an angle that corresponds to the first minimum of the receiver polar diagram. The exact angle of tilt of the receiver crystal necessary to obtain the maximum attenuation of the third-time-round signals is found by trial and error.

The quartz crystals have a natural frequency of 10 Mc/s. Owing to the attenuation of the ultrasonic waves in mercury, the frequency characteristic of the delay line peaks at about 7 Mc/s, and this is chosen as the frequency of the carrier wave. The signals are modulated on this carrier and fed to the transmitting crystal via a power amplifier. At the receiver the signals from the crystal pass via an R.F. amplifier and detector to a video amplifier. From here they pass to the channel-separating circuits (*fig. 14a*).

Channel-separating circuits

There are three sets of pulses passing simultaneously through the delay line: particle intercept pulses, "count" pulses and synchronizing pulses for the flying-spot time-base (fig. 14b). These are identified as follows. The "count" pulses are always superimposed on the intercept signals, as we have seen; selection between these signals can therefore be achieved by amplitude discrimination. The synchronizing pulses are separated in time from the other signals in that they occur during a period in which the line scan is blanked off. The blanking pulse for this purpose is obtained from two monostable flip-flops operated by a Schmitt trigger circuit which forms part of the line time-base and synchronizing circuits; see next section. This blanking

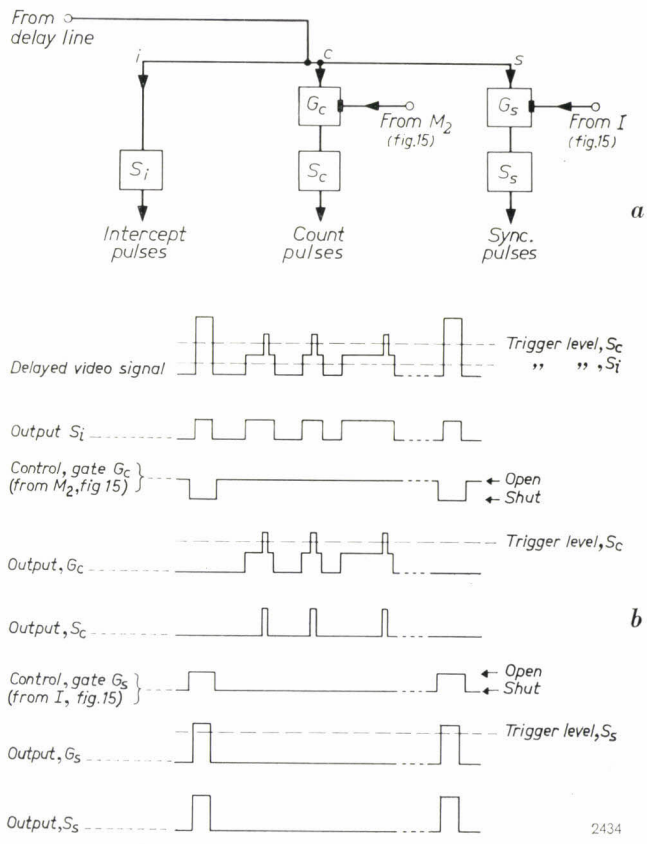


Fig. 14. *a*) Block diagram of channel-separating circuits. *i* intercept channel, *c* count-pulse channel, *s* sync-pulse channel. *G* are gates, and *S* Schmitt trigger circuits.
b) Waveforms in the three channels. In channel *i*, besides the delayed intercept signals, there are sync pulses. The fact that sync pulses are also fed to the clamp of the count-pulse generator does not matter because the scanner is blanked off when the sync pulses arrive. Channel *c* is blocked to sync pulses by the scanner blanking pulse which shuts gate G_c. Channel *s* is blocked to count pulses by the inverse of the scanner blanking waveform: gate G_s is thereby shut during the active scanning period.

pulse also shuts gate G_c in fig. 14*a* during the interval indicated in fig. 14*b*. This means that no sync pulses can enter the count-pulse channel (No gate is necessary in the intercept-signal channel as the sync pulses are present only during the blanked-off period, when there are no intercept signals; their arrival at the clamp in channel *1* (fig. 7) is then of no consequence.) The inverse of the blanking pulse is used to shut the gate G_s during the *active* scanning period (~700 μsec), so that no count pulses can enter the sync channel.

This completes the measures necessary to separate the signals of the three channels.

The sync-pulse amplitude is slightly greater than that of the count pulses and represents in fact the peak carrier amplitude; this is used as a datum level for the automatic gain control in the R.F. amplifier. ("peak white" in television systems with positive modulation).

Time-bases, synchronization and blanking waveforms

The line time-base is a Miller integrator linear-ramp generator, in which scan and flybaek voltages are initiated by a Schmitt circuit. The free-running time of this arrangement is made somewhat greater than the required total line period and the Schmitt circuit is triggered by its own output pulses delayed in delays *1* and *2* (fig. 7). In this way the line time-base is synchronized with the actual total delay, so that variations in the delays (especially in the mercury delay line, which is temperature-dependent) cannot affect the phasing between direct and delayed signals.

The frame-scan time-base is also a Miller integrator triggered by a Schmitt circuit. The frame period is here determined by the free-running time of the circuit or, in later versions of the instrument, by a counting circuit which locks the frame period to a specified number of lines (see note [17]).

The waveforms for the scanner blanking (spot suppression) and for the channel separator circuits are produced by two monostable flip-flops in cascade The first of these (M_1, *fig. 15a*) is fed with the flyback pulses from the Schmitt circuit of the line time-base. The leading edge of these pulses triggers M_1, which returns to its original state after an interval of about 250 μsec. Differentiation produces a negative pulse which thus lags the end of the flyback by about 50 μsec. This sharp negative pulse triggers the second flip-flop M_2, which has a relaxation time of about 700 μsec. As may be seen from fig. 15*b*, the output of M_2 gives square negative pulses of about 300 μsec duration, separated by intervals of about 700 μsec (their total is of course equal to 1010 μsec).

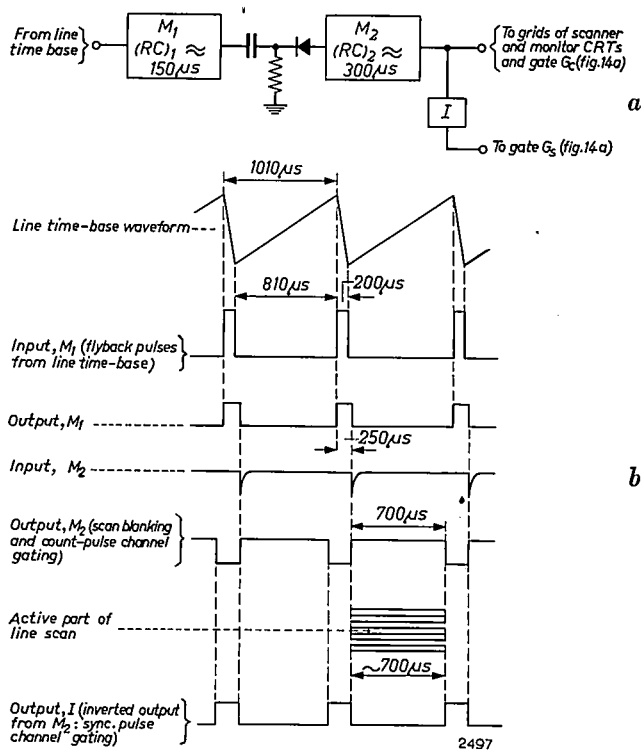


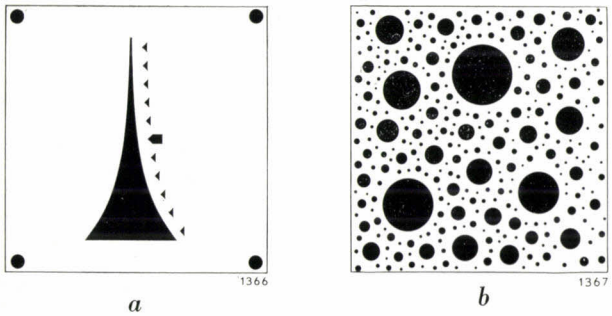
Fig. 15. *a*) Two monostable flip-flops connected in cascade are used to produce the scan-blanking and sync-strobe waveforms, as shown in (*b*).

These pulses are applied to the grids of the scanning and monitor tubes to suppress the spot during flyback and for a small portion of the line period at either extremity. The pulses are also applied to the count-pulse gate in the channel-separating circuits (G_c in fig. 14*a*), so that the sync pulses are denied entry to the count-pulse channel.

The inverse of the above waveform is applied to the sync-pulse gate in the channel-separating circuits (G_s in fig. 14*a*). The pulses shut gate G_s during the active scanning period, thus preventing count pulses from entering the sync channel.

Calibration and performance

The eleven settings of the sizing potentiometer correspond to *nominal* sizes from 0.20 to 6.40 mm. In order to have a check on the absolute sizes, and also to facilitate initial adjustment of the instrument, it is desirable to use a calibration sample such as the wedge shown in *fig. 16a*. This is easily prepared by making a drawing of known dimensions, and reducing it photographically on to 35 mm film. With the sizing potentiometer on the central size (1.14 mm, bold arrow in fig. 16*a*), the clipping level of the video pulse shaper (fig. 7) is adjusted till the displayed image at this point is just tagged (cf. fig. 6), i.e. just counted. The sizing potentiometer is



a *b*

Fig. 16. *a*) Calibration wedge, the width opposite each arrow increasing by a factor $\sqrt{2}$ between one arrow and the next. *b*) Test sample of circular spots whose size distribution is known. Both (*a*) and (*b*) represent a field of 20 × 20 mm on the actual negative. (For these samples and the calibration procedure described in the text we are indebted to Mr. W. A. Welling, Royal Dutch Shell Laboratories, Amsterdam.)

then set at the smallest range (0.2 mm) and the photomultiplier gain adjusted until only the image above the first arrow is tagged in the display. The adjustment may be repeated if necessary to obtain the optimum settings.

Whilst the absolute accuracy of the sizing and counting functions of the instrument are dependent on the shape factor of the particles being analysed and on the sharpness and contrast of their photographic image, the repeatability of results obtained from samples of the same material is of a very high order. For particles of circular profile it is found that, due to the line width (i.e. spot diameter) and spacing, about 5% of those lying in the smallest size range (0.2-0.28 mm) will be missed. Repeatability can be checked by means of a test sample such as that in fig. 16*b*, representing a known size distribution.

Table I shows some results of a series of tests carried out with such a test sample (not the one illustrated) on an instrument which had been in service for one year. These results illustrate the repeatability of results when the instrument parameters are re-adjusted between consecutive analyses of the same test sample.

Fig. 17 shows a photograph of spray droplets formed by the break-up of a jet from a fan-spray nozzle. The droplet size analysis of such a photograph is shown in *fig. 18*; two sets of points are plotted, representing size distributions measured at

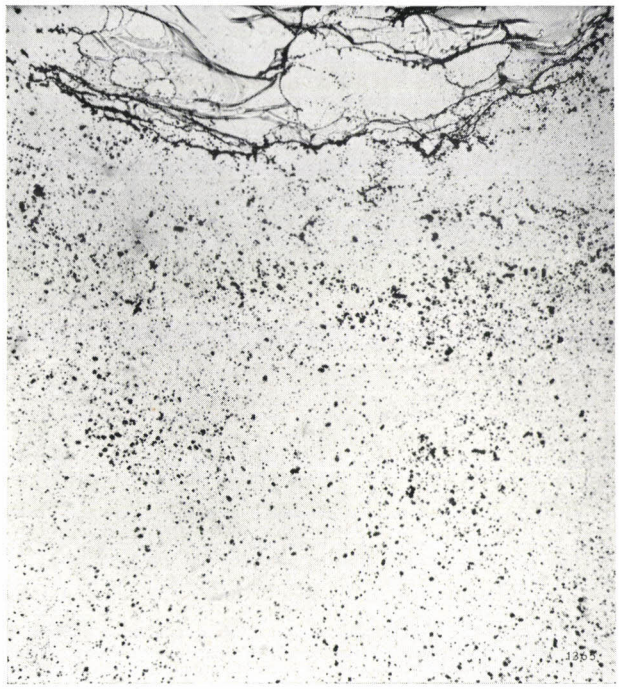


Fig. 17. Spray droplet photograph taken by a spark-photographic technique during the development of an atomizing system for liquid fuels. As reproduced here, the droplet spray appears at about actual size. The original negative scanned in the analyser had a magnification of 2× actual size.

Table I. Total count and size distribution of "particles" in an artificial test sample of known size distribution, as measured under various conditions. Each count represents the mean of 10 complete scans of the field.

Condition	Total count	Sizing: number of particles greater than d (mm) =										
		0.10	0.14	0.20	0.28	0.40	0.56	0.80	1.12	1.60	2.24	3.2
Actual number of particles in test sample	426	426	252	147	82	38	19	11	6	3	1	0
Initial state	426.0	400.2	250.9	143.9	82.0	38.0	19.0	11.0	6.0	3.0	1.0	—
20 minutes after switch-on	426.0	400.0	251.9	145.7	81.0	38.0	19.0	11.0	6.0	3.0	1.0	—
Photomultiplier gain changed slightly	425.9	388.9	243.0	142.9	80.6	38.0	19.0	11.0	6.0	3.0	1.0	—
Photomultiplier gain restored to initial value	426.0	390.1	245.0	144.6	81.1	38.0	19.0	11.0	6.0	3.0	1.0	—
Clipping level adjusted slightly	426.0	398.4	252.0	146.9	81.6	38.0	19.0	11.0	6.0	3.0	1.0	—

two slightly different positions on the same photograph. The fact that both sets of points lie almost on the same curve implies not only a substantially unvarying droplet size distribution but also a high degree of stability and reproducibility in the instrument.

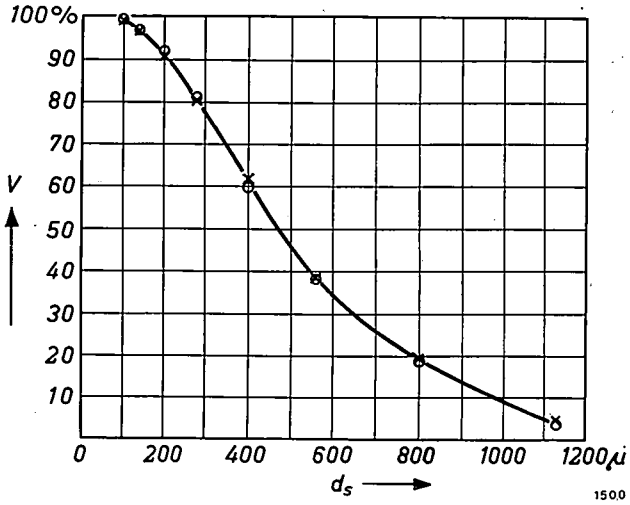


Fig. 18. Plot of two series of measurements of the size distribution of droplets from a fan spray nozzle using water at 25 pounds/inch². × Test A, sample field about 6 cm from nozzle. o Test B, sample field about 7 cm from nozzle. The counts are here plotted after conversion from a diameter size distribution to a volume distribution: V represents the total volume of all particles having diameters greater than d_s. (Photograph (fig. 17) and measurements by courtesy of Dr. J. D. Lewis, R. P. E., Ministry of Aviation, London.)

Summary. The problems involved in the automatic counting and sizing of particles are outlined in an introductory section to this article. Various possible scanning systems are described, together with some of the methods that have been used to avoid or allow for the effects of multiple or lumped counts. It is shown that flying-spot scanning together with some form of line-to-line signal association represent a very effective solution to the problem. For practical reasons line-to-line signal association can best be achieved by use of a line-to-line memory rather than by double-spot or guard-spot systems. Sizing can be conveniently performed by logical circuits that reject pulses shorter than a certain pre-set length. The following sections of the article describe the Mullard particle analyser. The sample is in the form of a 35 mm photograph (negative or positive) made from the original sample by direct photography, photo-micrography or electron microscopy. This greatly simplifies the optical arrangements and facilitates initial adjustments. With the aid of block diagrams and waveforms it is shown in detail how the instrument makes total counts and measures size distributions. The operation of some of the sub-units is explained with the aid of simplified circuit diagrams. Among the sub-units dealt with are the count-pulse generator, the sizing discriminator, the (mercury) delay line and the channel-separating circuits. The resolution is discussed in relation to the scanning raster and bandwidth requirements. The article concludes with a short section describing initial adjustments and calibration with artificial samples, and giving some results which demonstrate the reproducibility of the instrument.

A GAS-DISCHARGE INDICATOR TUBE FOR TRANSISTORIZED DECADE COUNTING CIRCUITS

by T. P. J. BOTDEN.

621.387:621.374.32

The problem of indication in transistorized decade counting circuits has long awaited a really satisfactory solution. The article below describes a gas-discharge decade indicator tube which responds to the small signals delivered by transistor circuits.

A scaling circuit, irrespective of whether it forms part of equipment for monitoring radioactivity or is a component of other counting equipment, obviously depends for its usefulness on being able to display in some way or another the result of a count.

In the case of equipment for monitoring radioactivity, for example, the pulses delivered by the radiation detector (e.g. a Geiger-Müller tube or a scintillation counter) are fed after amplification to a sequence of scaling circuits. With scale-of-two stages, the last stage is often followed by a mechanical register (which can handle no more than about 10 pulses per second); the state of the individual scaling stages can be seen from a neon lamp — lit or unlit — in each stage. Decade scaling circuits are usually equipped either with indicating neons — 10 per stage — or each stage consists of a decade scaler tube, which performs the dual function of scaling and indication — each tube thus indicating one digit of the number counted. Such scaler tubes may be either of the gas-filled type (decatron ring-electrode scalers) or of the vacuum type (decade scaler tube E1T).

Apart from giving a visual indication, as in a

mechanical register, scaler tubes thus form the primary functional elements of the circuit [1]). This means that the resolution of the first decade — which is the critical one — is determined by the tube; for the gas-filled tube this is about 300 μsec, and for the E1T tube about 30 μsec. The gas-filled type can therefore only be used where the highest counting rates are not required. In counting circuits from which an exceptionally high resolution is demanded, e.g. 0.1 to 1 μsec, the functions of scaling and indicating are again separated. The scaling is performed by vacuum-tube circuits (having a resolution down to 0.1 μsec) or perhaps by a "trochotron", a multi-electrode tube having a resolution of about 1 μsec [2]); the indication is provided by neon lamps or by numerical indicator tubes.

A resolving power almost as high as obtained with the best electron-tube circuits can be achieved with transistor circuits, which have the advantages of very compact construction and low energy consumption. Hitherto, however, the means of indication had presented difficulties inasmuch as a transistor circuit is unable to deliver a signal of sufficient amplitude to ignite a neon lamp or numerical indicator tube [3]). Moreover, a circuit of this type cannot provide the current needed to produce a clearly visible discharge without jeopardizing the counting operation. Further, in the case of decade scalers it is more convenient to have only one indicator tube for each decade. In this article we shall discuss a gas-discharge tube which provides an answer to these problems [4]). This tube, which is purely an indicator and does no counting, has an anode and ten cathodes arranged in a circle; the gas discharge can be shifted as required from one cathode to another by a signal of only 5 volts and less than 50 μA. The tube is not yet commerciably available.

Principle of the indicator tube

The structure of the tube is shown schematically in *fig. 1*. The face of a molybdenum ring K is coated with ten sectors of an insulation material (shaded in the figure). The ten sectors in between them each act as the cathode. About 5 mm above this ring there is a second ring A, of nickel wire, which acts as the anode. Above each of the ten cathodes can be seen a wire electrode T, the "trigger", by means of which the discharge is initiated at the desired place. The distance between the triggers and the cathodes is about 1 mm. *Fig. 2* illustrates some of the arrangements of cathodes and triggers; a photograph of a complete electrode system is shown in *fig. 3*.

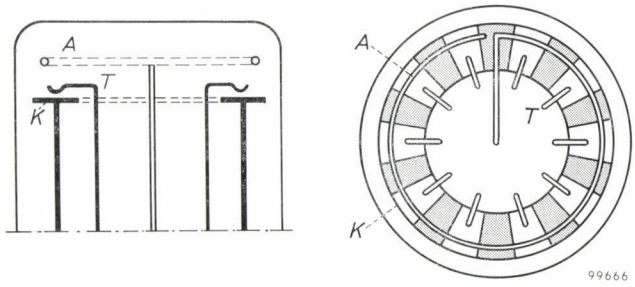


Fig. 1. Schematic representation of the indicator-tube electrode system. The surface of a molybdenum ring K is divided into twenty sectors. Ten are coated with a layer of insulating material (shaded), the other ten function individually as cathodes. The circular nickel wire A is the anode. Between the anode and each of the ten cathodes is situated an auxiliary electrode T, which functions as trigger electrode.

The tube is filled with neon gas to which 0.1% of argon has been added; the gas pressure is about 15 cm Hg. In order to obtain a clean cathode surface,

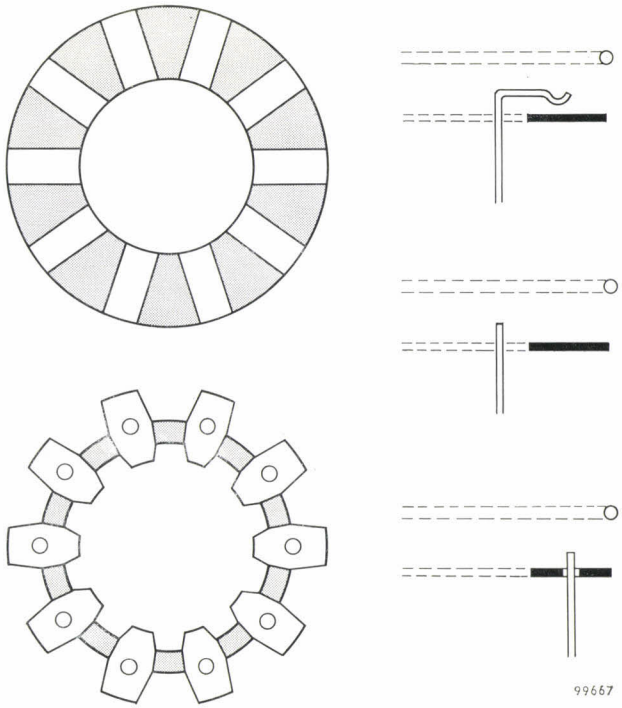


Fig. 2. Various arrangements of cathodes and triggers.

[1]) In ring-electrode scalers each count pulse causes a gaseous discharge between the disc-shaped anode and one of the ten cathodes arranged around it. The next pulse shifts the discharge to the adjacent cathode, and so on. Every time the discharge passes cathode number "zero", a pulse is delivered to the tube in the following scaling stage. (See e.g. R. C. Bacon and J. R. Pollard, Electronic Engng. **22**, 173, 1950 and G. H. Hough and D. S. Ridler, Electr. Commun. **27**, 214, 1950.)

The operation of the decade scaler tube type E 1 T has been described in this journal (A. J. W. M. van Overbeek, J. L. H. Jonker and K. Rodenhuis, Philips tech. Rev. **14**, 313-326, 1952/53).

[2]) See e.g. H. Alfven and L. Lindberg, Acta polytech. (El. Engng.) **2**, 106, 1949.

[3]) Cf. H. F. Stoddart, Nucleonics **17**, No. 6, 78, June 1959.

[4]) The development of this tube stemmed from an idea of M. van Tol of this laboratory.

the cathode is sputtered during manufacture [5]); the sputtered material on the glass wall helps to keep the gas uncontaminated. The reason for adding argon will be explained below.

The tube is fed with a rectified alternating voltage, which is not smoothed. A discharge is initiated when the amplitude of this voltage is sufficiently high. For a half-wave rectified supply, the supply voltage rises

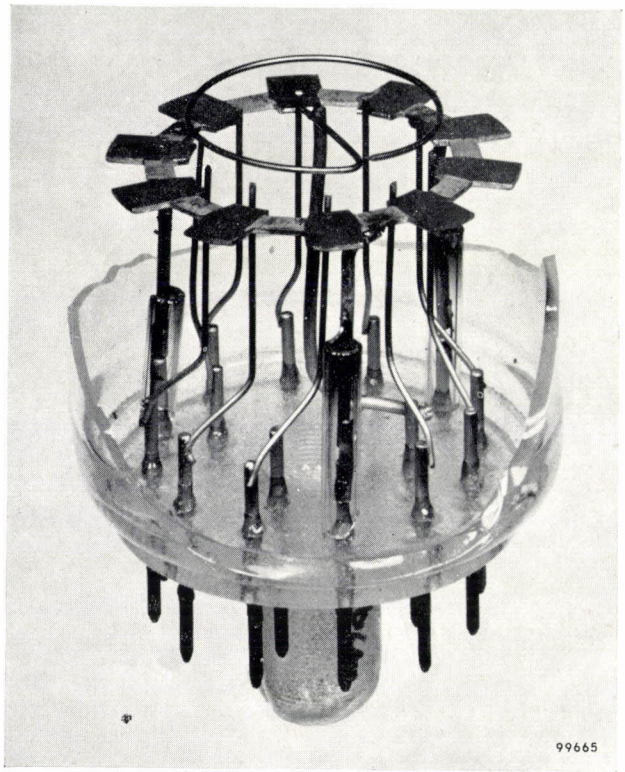


Fig. 3. An assembled electrode system.

to a maximum and drops to zero once in every mains cycle; the tube is therefore ignited and extinguished once per cycle. (However, a full-wave rectified voltage may also be used.) As can be seen from the circuit diagram in *fig. 4* (disregarding the blocks X_i for the moment), the triggers are at the same potential as the anode so long as there is no discharge. A discharge between the cathode and one of the triggers has a lower ignition potential than a discharge between cathode and anode, and therefore when the voltage begins to rise from zero a discharge first occurs between the cathode and one of the triggers. If the current produced by this auxiliary discharge is high enough — we shall return to this presently — the anode takes over the discharge almost immediately. The potential difference between cathode and anode then drops to the burning potential of the main discharge now occurring between these electrodes (a glow discharge), so that for the rest of that particular half cycle none of the other triggers can reach their breakdown potential. This process is repeated in every cycle of the mains frequency.

The place where the auxiliary discharge occurs can be selected by making the potential of the relevant trigger higher than that of the other triggers (and of the anode) by a small amount X. As a result this trigger reaches the breakdown potential earlier than the others and the discharge always recurs at the same position. If the voltage X is transferred to another trigger, the reignition in the next mains cycle will take place at that trigger, and so on. The periodic extinction of the discharge is thus essential in order to be able to displace the discharge from one position to another. It follows from the above that the tube can be "driven" with a signal whose amplitude is much smaller than the breakdown voltage itself. For the tubes made and tested by us, the amplitude of this signal was always less than 5 V.

A signal as small as this can readily be supplied by a transistor circuit. If the circuit is so designed that a signal X is applied to the trigger T_1 for a count of 1, to the adjacent trigger T_2 for a count 2, and so on, one can read from the tube the total result of a count. Evidently, it is immaterial whether or not the tube can follow a rapid counting operation because upon the next reignition after the completion of the counting operation the tube always burns at the position corresponding to the final result of the count. Since the power for the main discharge is not drawn from the transistor circuit, it is easy to ensure that this discharge will be sufficiently bright to provide a clear visual indication.

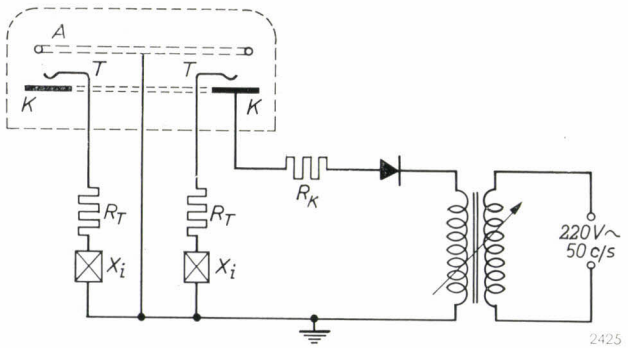


Fig. 4. Circuit diagram of indicator tube. A anode (earthed). K cathodes. T triggers. X_i voltage source providing the control signal. The current of the main discharge is limited by the resistor R_K, that of the auxiliary discharge by R_K together with one of the resistors R_T. The tube is supplied with a rectified alternating voltage (unsmoothed).

[5]) In this process the cathode is exposed for some time to intense ionic bombardment. Cathode material is thereby sputtered onto the inside wall of the glass envelope, where it performs the dual function of gettering and of preventing gas molecules in or on the glass wall from escaping. See e.g. T. Jurriaanse, Philips tech. Rev. **8**, 272, 1946.

As can be seen in fig. 4, the anode is earthed. This makes it possible to earth one of the two terminals of the voltage sources that supply the control signal (represented by X_i in fig. 4). In the practical application of the tube these sources are part of a transistor circuit.

Closer consideration of the gas discharge and the method of ignition

It has been shown in the foregoing that the operation of the tube depends, among other things, on the fact that a gas discharge can be initiated by a voltage lower than the normal breakdown voltage (but not, of course, lower than the burning voltage), and that this is done by producing an auxiliary discharge between the cathode and a trigger electrode. The extent to which the ignition potential of the main discharge is reduced depends on the current I_T of the auxiliary discharge. This is illustrated in *fig. 5* for one of the ten cathode-trigger positions and for a tube in which the pressure of the neon-argon filling was 14 cm Hg. It is seen that even for a trigger current I_T of only a few tens of μA the anode voltage V_a at which the main discharge is initiated is already much lower than the breakdown voltage measured in the absence of an auxiliary discharge. For maintaining the auxiliary discharge the current drawn from the transistor circuit need therefore only be, say, 50 μA (the figure mentioned earlier), which does not affect the operation of the transistor circuit.

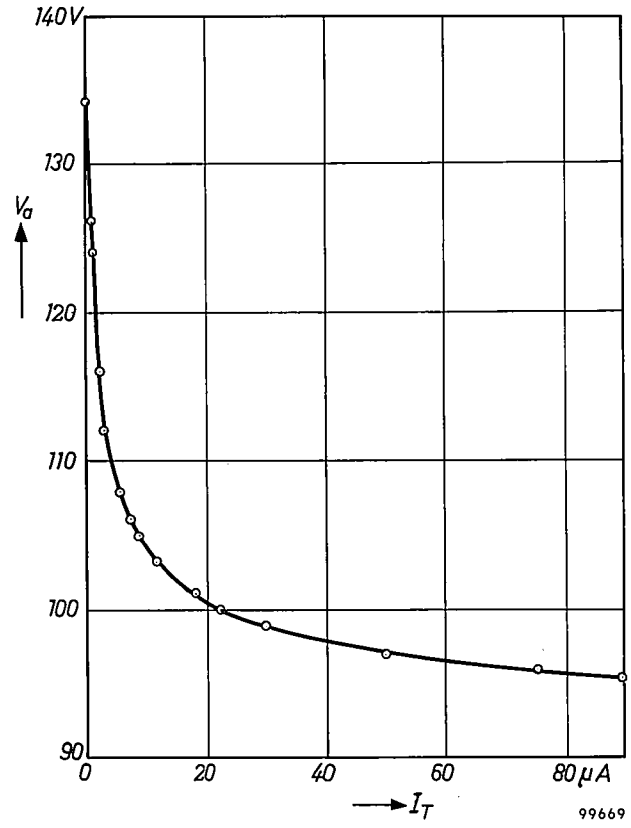


Fig. 5. Ignition potential V_a of the main discharge as a function of the current I_T of the auxiliary discharge for one of the ten cathode-trigger positions. The pressure of the gas filling (Ne + 0.1% Ar) was 14 cm Hg.

There are two reasons for the lowering of the ignition potential. In the first place, charge carriers (positive and negative) diffuse from the auxiliary discharge to the space between trigger and anode. Consequently, the current I_0 that begins to flow between cathode and anode when the supply voltage rises from zero is already, before ignition, greater than at the other cathodes. If this current exceeds a certain minimum value, the ignition voltage is no longer independent of it but decreases according as I_0 increases [6]). In the second place, the auxiliary discharge gives rise to a certain space charge (of positive ions) at the cathode. As a result the electrons in the space immediately in front of this cathode encounter more favourable conditions (a higher electric field) than those at the other cathodes for acquiring sufficient energy to ionize the gas atoms. At the other cathodes the field strength is determined solely by the applied voltage and the geometrical configuration [7]).

It should be noted here that the auxiliary discharge is neither a Townsend nor a glow discharge; the values of current and burning voltage are found in the descending transition region of the discharge characteristic between the two horizontal portions corresponding to the Townsend and glow discharge regimes (cf. *fig. 6*).

Factors determining the minimum value of the control voltage

If the potential difference V_d at which breakdown occurs between the trigger and cathode always had the same unvarying value for all trigger-cathode positions, the amplitude of the control signal could be extremely small. After all, this need not be much greater than the maximum difference found between the V_d values for the ten triggers. We shall now consider the factors that determine the actually occurring differences in V_d, and the measures adopted in the indicator tube to reduce them to a value lower than the 5 volts referred to above. First we shall recall the way in which a discharge is

---

[6]) For a brief discussion of the mechanism of this effect, see C. H. Tosswill, Cold-cathode trigger tubes, Philips tech. Rev. **18**, 128-141, 1956/57. See also M. J. Druyvesteyn and F. M. Penning, Rev. mod. Phys. **12**, 87-174, 1940.

[7]) See e.g. F. M. Penning, Electrical discharges in gases, Philips Technical Library 1957, Chapter VIII, or the article by Druyvesteyn and Penning quoted under [6]).

brought about, and the conditions to be fulfilled if the discharge is to be self-sustaining.

If a not unduly high voltage is applied across two electrodes in a gas-filled space, a minute current will

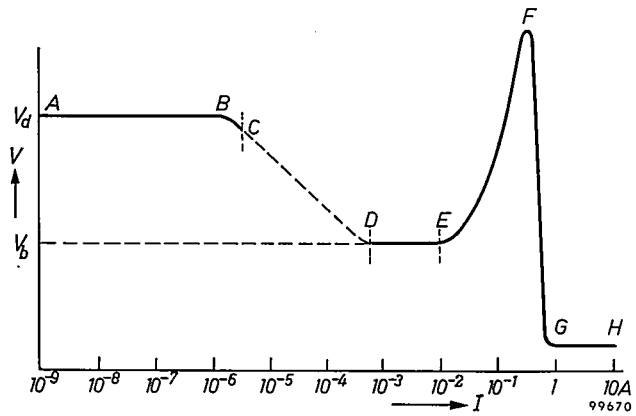


Fig. 6. Current-voltage characteristic of a self-sustained gas discharge. In the region AC of the Townsend discharge no space charge (of positive ions) has yet formed in the region AB, and the burning voltage is equal to the breakdown voltage V_d. A space charge begins to form in the region BC, and with rising current this continues into the region beyond C until at D the region of the normal glow discharge is reached (DE). Here too the burning voltage (which is now much lower than the breakdown voltage) is independent of I. The current density is likewise independent of I; as the current increases, a larger part of the cathode surface is covered, until at E the whole cathode is covered. A further rise in current causes the current density and the burning voltage to increase (anomalous glow discharge; EF). The region FGH is the arc-discharge region.

flow whose magnitude is determined by the number of charge carriers created per second by some external cause, such as cosmic radiation or the photo-electric effect of daylight. (If the tube is shielded from all radiation, the current drops to zero.) As the voltage is raised the current at first remains unchanged, then gradually increases and finally increases very rapidly, i.e. breakdown occurs. The gaseous discharge thus produced cannot be extinguished by suppressing the external cause that gives rise to the presence of the initial charge carriers; the production of charge carriers by the discharge itself—which is also the cause of the gradual increase of the current with rising voltage before the breakdown occurs — is now great enough to maintain the discharge. The charge carriers are generated both as a result of gas atoms being ionized by collision with electrons moving towards the anode (or trigger), and by ions reaching the cathode and liberating electrons from its surface. The discharge can remain self-sustaining if as many of these electrons are liberated from the cathode per second as are lost from the gas to the anode. The potential difference then required between cathode and anode (trigger) is equal to the breakdown voltage V_d (in the absence of any space charge).

If a voltage slightly higher than V_d is applied across the electrodes, nothing will happen until there is a free electron present at a suitable position in the space between the electrodes. Once such an electron exists, it will of course move through the electric field towards the anode. In doing so it ionizes gas atoms, as a result of which electrons are released which ionize other gas atoms, and so on. Because of the random character of the processes taking place in the gas and on the cathode, however, the "avalanche" thus produced may sometimes break off pending the arrival of a new initial electron. Not until the number of ions and electrons generated is extremely high is the chance of such an interruption eliminated. The time elapsing between the application of the voltage and the attainment of this situation is called the statistical delay. Since we are concerned here with random processes, this delay time may, in principle, have any value. The time elapsing subsequent to the formation of this stable avalanche up to the moment that the avalanche effect ceases to increase in strength is the "build-up" time of the discharge. At the beginning of this time interval there are already large numbers of ions and electrons present, and therefore the statistical fluctuations in the build-up time are extremely small and cannot be measured: the build-up time can thus be regarded as constant.

The condition that the discharge should supply its own electron requirements can be expressed by the equation:

$$\gamma(e^{\eta V_d} - 1) = 1. \quad \ldots\ldots \quad (1)$$

Here γ is the average number of electrons released from the cathode by one gas ion, and η is the average number of electrons produced by each electron in the gas per volt of traversed potential difference. The value of γ depends on the velocity and kind of the ion and on the condition of the cathode surface. From (1) it follows that V_d depends on γ and on η.

Applying this to the new indicator tube, we may conclude that the differences in the γ value of the various cathodes are due solely to the condition of the cathode surfaces; the gas filling is of course the same in all cases. The surfaces of the ten cathodes are cleaned during manufacture by the sputtering process referred to, thus considerably reducing the differences in the γ value. The gettering action of the sputtered molybdenum on the glass wall helps to keep the gas clean and thus maintains this favourable situation over a very long time. Another fortunate circumstance is that V_d does not depend very markedly on γ. This can be understood from

the following reasoning. In the left-half of equation (1) the exponential term is $\gg 1$. To a good approximation we can thus reduce (1) to

$$V_d = -\frac{1}{\eta} \ln \gamma. \quad \ldots\ldots \quad (2)$$

The breakdown voltage, then, varies in proportion to $\ln \gamma$. Since $\ln \gamma$ varies only slowly with γ, the minor differences in γ that still exist after the cathode has been sputtered have scarcely any effect on V_d.

It also appears from (2) that V_d depends more strongly on η than on γ. Now η is by no means constant; its value depends on the ratio of the electric field strength E and the gas pressure p. For a given electrode configuration, E is inversely proportional, at a given voltage, to the linear dimensions of the electrodes, e.g. for two flat plates E is inversely proportional to the distance d between them. In this case, then, V_d is a function of pd (Paschen's law). *Fig. 7* demonstrates this relation for the breakdown of various gases between flat molybdenum electrodes [8]). It can be seen that V_d shows relatively little variation in a wide range of pd values in the case of a gas mixture of neon and 0.1% argon.

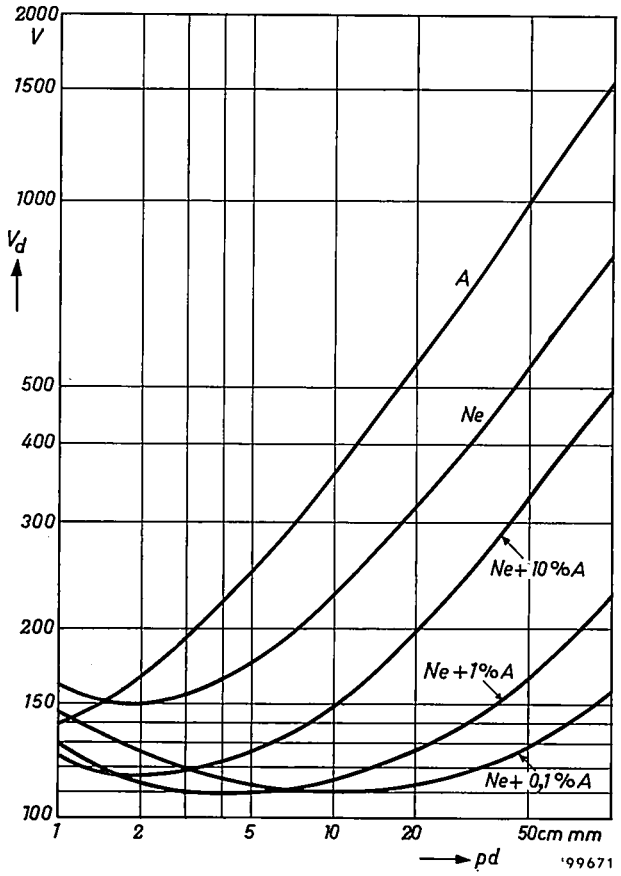


Fig. 7. Breakdown voltage V_d of various gases as a function of the product pd, using flat molybdenum electrodes (Paschen curves). For a mixture of neon and 0.1% argon there is relatively little variation of V_d over a wide range of pd.

[8]) S. M. Frouws, 3° Congr. Int. Fenomeni d'ionizzazione nei gas, Venice 1957, page 341.

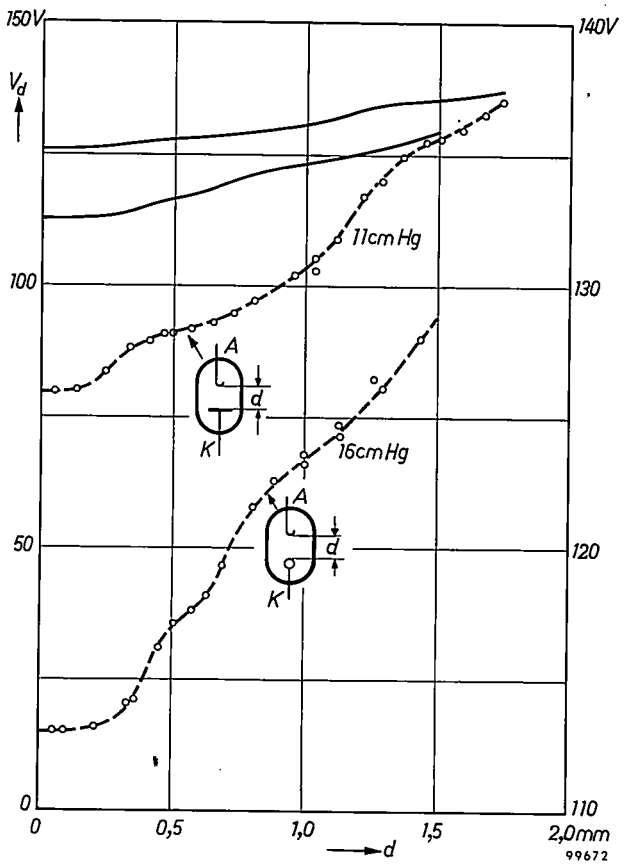


Fig. 8. Breakdown voltage V_d of a trigger-cathode combination as a function of the distance d between these electrodes (solid curves). The solid curve at the top applies to a flat molybdenum cathode, the lower one to a molybdenum-wire cathode. In the case of the flat cathode, V_d varies only by about 0.8 V per tenth of a millimetre for $d \approx 1$ mm. The gas pressure for the upper curve was 11 cm Hg, and for the other 16 cm Hg. The curves are derived from two series of measurements which are plotted as the dashed curves (V_d scale on right of figure).

Although the curves shown in fig. 7 do not entirely apply to the new indicator tube with its wire-shaped triggers, it was to be expected that here too, owing to the use of the Ne-Ar mixture, the ignition potential of the auxiliary discharge would not vary much with the electrode spacing. *Fig. 8* shows the results of an investigation into this question. In one series of measurements the trigger was mounted opposite to a flat molybdenum cathode, and in the other series opposite to a molybdenum-wire cathode. It can be seen that V_d does not in fact vary to any marked extent with d, and at very small d values it is virtually constant. With the flat cathode and $d \approx 1$ mm, V_d varies by about 0.8 V when d is increased or decreased by one tenth of a millimetre, i.e. by 10%. Thus, the choice of gas filling also helps to keep down the relative difference in breakdown voltage between the various trigger-cathode spaces.

Besides the condition of the cathode surface and the geometry of the electrodes there is a third factor affecting the minimum amplitude of the control signal. This is the phenomenon that, during operation of a given tube, the breakdown voltage V_d of any particular cathode-trigger position is not entirely constant, but shows certain fluctuations. As these are related to one of the delay effects mentioned above, and thus the magnitude of the fluctuations in V_d depends on the rate at which the trigger voltage rises, we shall defer the discussion of this subject to the following section, which deals with the relation between the minimum required signal amplitude and the properties (amplitude, frequency, waveform) of the supply voltage.

Voltage supply

As we have seen, for the purpose of shifting the discharge readily from one point to another the indicator tube is fed with a half-wave rectified

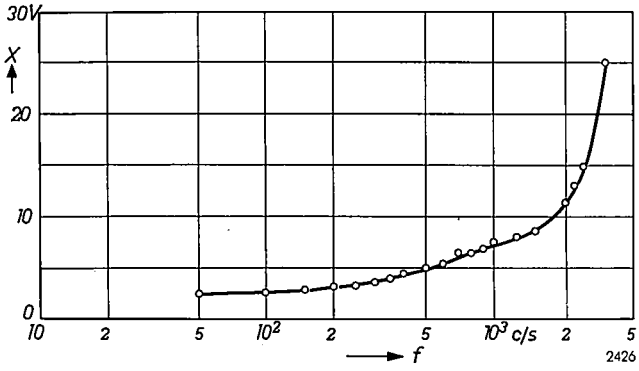


Fig. 9. The minimum required control voltage X as a function of the frequency f of the supply voltage. At frequencies above 3000 c/s, X is higher than 25 V (the difference between ignition and burning voltage) and therefore the discharge can no longer be displaced.

alternating voltage. For a visual indication the only requirement to be made of the frequency of this voltage is that the glow should not flicker. A frequency of 50 c/s is quite high enough for this purpose.

Investigation of the way in which the minimum required control voltage depends on the frequency of the supply voltage shows (see *fig. 9*) that the control voltage is substantially constant at frequencies up to about 300 c/s, but increases at frequencies higher than this. At frequencies above about 3000 c/s the discharge can no longer be displaced. The main reason for this is that, after the gas discharge is extinguished, all the ions and electrons constituting the plasma do not immediately disappear. If a voltage is again applied while there is still some residual plasma present, the ignition

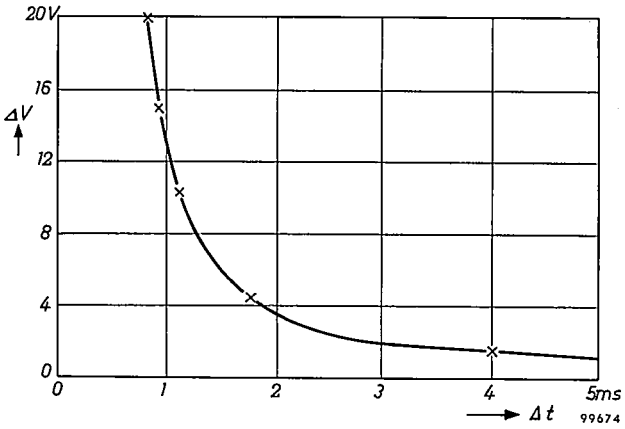


Fig. 10. The decrease ΔV of the ignition voltage (from the ignition voltage required when the tube has not been ignited for a long time) as a function of the time interval Δt between the moment a discharge is initiated and the moment at which the preceding discharge was extinguished. The smaller the value of Δt, the greater is ΔV. This is due to the fact that not all ions and electrons disappear immediately after extinction of the discharge.

takes place at a lower potential than otherwise. The drop is greater the greater the residual plasma density, that is to say, the shorter the time Δt between the extinction of the discharge and the re-application of the voltage. If the drop is such that the breakdown potential has fallen to the level of the burning voltage, the tube will always ignite at the same place. *Fig. 10* shows the result of a series of measurements performed to determine the drop in breakdown voltage on an indicator tube of the type described here. It is seen that for $\Delta t = 1$ msec, the decrease ΔV in the breakdown voltage amounts to as much as 13 V, and increases rapidly for even shorter times. This roughly agrees with the measurements given in fig. 9. The difference between the normal ignition voltage and the burning voltage (about 25 V) corresponds in fig. 10 to a Δt of 0.7 msec; as regards the minimum trigger voltage required for displacement of the discharge, this corresponds to a maximum supply frequency of approximately 1000 c/s. In view of the fact that the experimental conditions in the measurements summarized in fig. 10 differed from those normally encountered, the agreement with the value derived from fig. 9 may be said to be satisfactory.

In order to obtain some quantitative information on the effect of delay phenomena, recordings were made of the current in the main and auxiliary discharges, and also of the total current in both discharges together, as a function of time after the application of the voltage. The circuit used for the recording is shown schematically in *fig. 11*. The objects of this investigation were to determine the values of the statistical delay, and also to discover the rate at which the auxiliary discharge builds up,

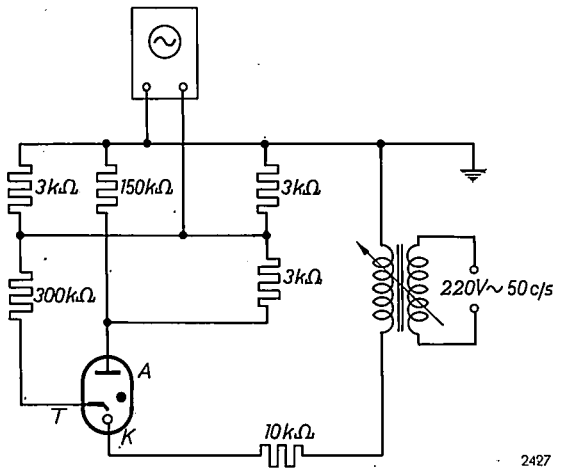


Fig. 11. Circuit for recording waveforms of the trigger current, the main current or the sum of certain selected fractions of each. The time-base period of the oscilloscope was chosen such that the full picture width corresponded to 30 μsec. A separate synchronizing circuit was dispensed with by triggering the time-base with the same mains voltage as that supplied to the indicator tube.

and the delay in the ignition of the main discharge.

Some of the oscillograms recorded are drawn in *fig. 12*. The curves in fig. 12*a* refer to the main-discharge current. It can be seen that they are all of the same form, indicating a uniform build-up time, but that they are somewhat displaced with respect to each other. This spread amounts to only a few microseconds. The extremes lie about 10 μsec apart. The curves in fig. 12*b* represent the sum of certain fractions of the currents of both discharges, obtained by suitable choice of the resistances in the circuit of fig. 11; the ratio of the fractions selected was thus not the same in both recordings. From this fact and from the appearance of the curves we may conclude that the first peak to occur is due to the auxiliary discharge, and the second to the main discharge. The ignition of the main discharge is seen to begin less than 5 μsec after the first peak. This "take-over time" and also the build-up time (cf. fig. 12*a*) are found to be virtually constant. The spread in the total delay time is therefore almost entirely due to the spread in the statistical delay.

From the result of the above experiments we may deduce that the effects involved here have no influence on the maximum permissible frequency of the supply voltage. The measured time intervals are roughly 100 times smaller than the deionization time.

Owing to the fact that the total delay shows a certain spread, this does, as we have seen, have some influence on the minimum required amplitude of the control signal. This is evident since the spread in the total delay also causes a spread in the breakdown voltage which is equal to the delay multiplied by the time derivative of the supply voltage at the value of the breakdown voltage (about 130 V). This time derivative is greater the greater the amplitude of the supply voltage, the frequency remaining constant, and also the higher the frequency is raised with the amplitude remaining constant. An idea of the magnitude of this effect is given by the following numerical example, using values encountered in practice. Taking a frequency of 50 c/s, an amplitude of 300 V and a difference in delay time of 10 μsec for two breakdowns, then the ignition voltages of these discharges differ in value by about 0.85 V; at an amplitude of 150 V they differ by about 0.35 V. Compared with a signal amplitude of 3 or 4 V, these amounts are not entirely negligible.

It may be inferred from the foregoing that the tube can only be fed with a square-wave voltage if the steepness of the square-wave edges is less than about 10^5 V/sec; otherwise, the minimum required control signal will be considerably increased.

In this connection it should be noted that the minimum control voltage is smaller the closer the

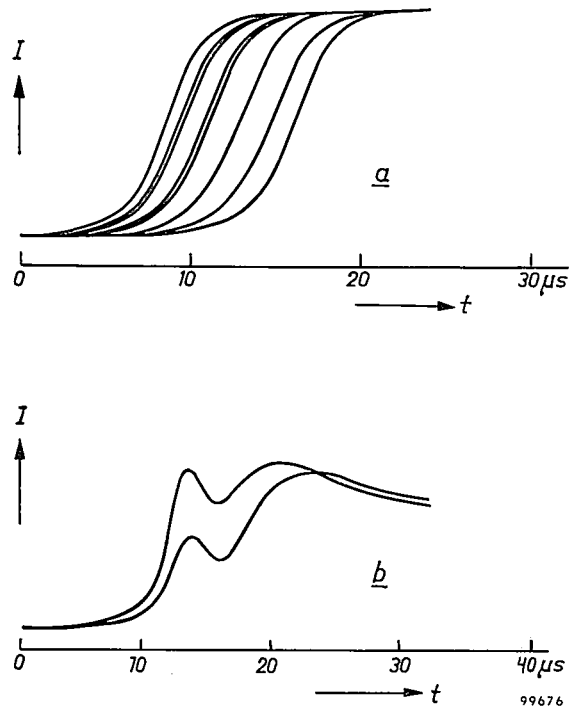


Fig. 12. *a*) Oscillograms of the main-discharge current. The curves are all of virtually the same form, indicating a constant build-up time, but there is a spread of some microseconds between the start of each curve (statistical delay). The outside curves lie less than 10 μsec apart.
b) Two oscillograms displaying the "summed" currents of auxiliary and main discharges. For both curves, the resistances in the circuit (fig. 11) are so chosen that the contribution of the auxiliary discharge current to the total signal is exaggerated with respect to that of the main-discharge current. In the one curve the ratio of the two contributions is different from the other curve. From the difference between the two it may be concluded that the first peak is due to the auxiliary discharge and the second to the main discharge. The main discharge is seen to ignite a few microseconds after the auxiliary current has reached its maximum.

electrode to which the discharge is to be transferred. In the light of what has been said in this section, this phenomenon is perfectly understandable. The electrons that initiate the avalanche in our tube originate, as a rule, from the discharge that took place during the preceding mains cycle. These electrons obviously diffuse faster and in greater numbers to nearby electrodes than to more distant ones. In regard to the discharge on these electrodes the statistical delay is therefore not only smaller but shows a smaller (absolute) spread, and hence gives a smaller spread in ignition potential.

It is due to this electron supply by diffusion that the statistical delay is as small as mentioned above. If the tube has been switched off for some time, so that the first electron must originate from another source, the delay may be appreciably greater. This does not, of course, prove troublesome in actual operation.

Visibility of the discharge

The visibility of a glow discharge depends on the nature of the gas and on the current density of the discharge. As already mentioned, the new indicator tube is filled chiefly with neon gas. From the viewpoint of light output neon is the best among the inert gases (other gases are chemically too reactive in gas discharges).

The current density is stepped up by choosing a fairly high gas pressure — the current density being roughly proportional to the square of the gas pressure — and also by producing an anomalous glow-discharge (cf. fig. 6).

Measurements have shown the brightness of the discharge to be about 1200 cd/m^2 (current density 10 mA/cm^2). This may be compared with the brightness of a 40 W fluorescent lamp, which is about 8000 cd/m^2. Partly because of its orange-red colour, the glow discharge is still clearly perceptible in normal daylight.

The anomalous discharge in the new tube is made possible by limiting the surface area of the cathodes; this also makes the indication easier to read. Without such limitation there would still be a glow discharge (in this case a normal one) near the trigger carrying the control signal, but it might take up quite an asymmetrical position in relation to that trigger.

Summarizing, it can be said that the new tube satisfactorily meets the existing need for a simple, easily readable decade indicator capable of being actuated by a low-energy signal of only a few volts. Tests have shown that the properties of the tube remain remarkably constant, both after intensive use and after long storage. The life of the tube, like that of other glow-discharge tubes, is expected to be very long.

———

Summary. For the read-out of transistor scaling circuits there is a need for indicator tubes (that do not themselves count) capable of being operated by low-energy signals of a few volts. The tube described has a flat annular cathode, the surface of which is divided into ten sectors, and a ring-shaped anode; the tube is filled with Ne + 0.1% Ar at a pressure of about 15 cm Hg. A gas discharge (anomalous glow discharge) is initiated at the desired place in the tube by means of one of ten auxiliary electrodes (triggers). The control signals supplied by the transistor circuit need only make the potential of the relevant trigger differ by 5 V from that of the (earthed) anode. The current of the auxiliary discharge required to initiate the main discharge is a mere 50 μA. The displacement of the main discharge to any desired position, corresponding to a given count, is made possible by the periodic extinction of the discharge. To this end the tube is fed with an unsmoothed, rectified alternating voltage, obtainable e.g. from the mains.

EXPERIMENTS WITH RADIOACTIVE PREPARATIONS OF THE ACARICIDE "TEDION V 18" *)

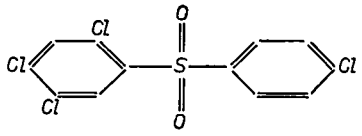
by J. HALBERSTADT **).

632.951.001.4:539.16

Under the proprietory name "Tedion V 18", Philips-Duphar are marketing preparations which destroy the eggs and larvae of mites, including the dreaded spider mite or "red spider" [1]). Trees treated with "Tedion" retain a residue which continues to exert its acaricidal action over a long period [2]). Tedion is not injurious to the tree, and is only very mildly toxic to warm-blooded animals.

In research on pesticides (the collective name for insecticides, acaricides, etc.) it is important to have quantitative data on the assimilation, transport and breakdown of these substances in the plant. In regard to their toxicity to warm-blooded animals, information is needed concerning the part the substances play in the metabolism of the animal (i.e. to what extent they are stored in the tissues and excreted from the body). An exact method of obtaining such data is provided by the use of radioactive tracers. The pesticide is "labelled" by partial replacement of one of its constituent elements by a radioactive isotope; the activity of the plant or animal treated with the pesticide is then measured to trace the distribution or location fo the tracer element.

This method has been adopted in the case of Tedion. The active ingredient of Tedion is 2,4,5,4′-tetrachlorodiphenyl sulphone:



By a process briefly described elsewhere [3]) this substance is labelled with the β-active isotope of sulphur, ^{35}S. The half-life of the isotope is 87 days. Preparations with this compound have been used in field experiments on apple trees and in experiments on the metabolism of rats. A brief account of both types of experiment is given below.

Experiments on apple trees

The experiments on apple trees were done with Tedion in two formulations, viz. as a *wettable powder* and as *miscible oil* (emulsifiable concentrate) [4]). Two grams of the powder (containing 20% by weight of Tedion) were suspended in two litres of water, and 5 millilitres of miscible oil (Tedion content 8%) were emulsified in two litres of water. In the grounds of the Boekesteyn Agrobiological Laboratory [5]) two apple trees of the "Golden Delicious" variety were carefully sprayed with each of these liquids. This was done at the beginning of July 1957, when the weather conditions were favourable. The specific radioactivity of the Tedion preparation was then 6.1 millicuries per gram.

The objects of the experiments about to be described were to determine the amount of Tedion that settled *on* the leaves, the amount that penetrated *into* the leaves, and the manner in which both amounts changed in the course of time.

As soon as the leaves had dried, two samples were taken of 20 leaves each. Like all subsequent samples, these were immediately subjected to the following treatment. First, the leaves were cut lengthwise in two. The 20 left halves were at once dried out for eight hours at 40 °C and then reduced to powder, after which the total radioactivity of this powder was measured. The 20 right halves were washed in nitromethane to remove the radioactive Tedion on the surface (nitromethane is the only known solvent for Tedion that does not attack the leaves). These half leaves were thereupon dried and pulverized like the others, and the radioactivity of the internal Tedion was then determined.

The radioactivity measurements were done on "infinitely thick" powder preparations in an experimental arrangement employing a Philips Geiger-Müller counter type 18 506, with end window, and a Philips counting unit, type PW 4035. The variation in the results of the measurements was about 3%.

To ascertain from the radioactivity the amount of Tedion taken up in the leaf samples, a comparison

*) Trade mark.

**) Now with the International Atomic Energy Agency at Vienna, formerly with N.V. Philips-Duphar. Weesp (The Netherlands).

[1]) J. Meltzer, Research on the control of animal pests, Philips tech. Rev. **17**, 146-152, 1955/56, in particular page 152.

[2]) J. Meltzer and F. C. Dietvorst, Action of Tedion on eggs and ovaries of spider mites, T. Planteziekten **64**, 104-110, 1958.

[3]) J. Halberstadt, Some experiments with radioactive preparations of 2,4,5,4′-tetrachlorodiphenyl sulphone, a new acaricide, Meded. Landbouwhogeschool en Opzoekingsstations van de Staat te Gent **23**, 788-794, 1958.

[4]) W. Duyfjes, The formulation of pesticides, Philips tech. Rev. **19**, 165-176, 1957/58.

[5]) Philips tech. Rev. **16**, 353, 1954/55.

sample was prepared. For this purpose 2.54 g of dry leaf powder, obtained from ten leaves that had not been sprayed, was mixed with a solution of 0.254 mg of radioactive Tedion, and again dried.

A comparison of the activities does not, however, yield directly the amount of Tedion but rather the amount of sulphur atoms introduced by the Tedion into the leaves. These belong only partly to chemically unaffected molecules of Tedion; the remainder of the sulphur has been incorporated in the plant in the form of glucosides or proteins, or is contained in the break-down products of the Tedion.

In order to discover what proportion of the sulphur was still present in free molecules of Tedion, the method of isotopic dilution analysis was used. A weighed amount of non-radioactive Tedion is added to the dried leaf material, and the Tedion is then extracted from the mixture with the aid of a suitable solvent. In this way the radioactive Tedion already present in the leaf powder is obtained in a solution together with the added non-radioactive Tedion. After concentration of the solution the Tedion largely crystallizes out. This preparation is purified by recrystallizing it until the melting point and the specific activity are constant. Let the latter be a and let that of the Tedion sprayed on the tree be a. If x milligrams is the original amount of radioactive Tedion present in the leaf powder, and g milligrams is the added amount of non-radioactive Tedion, then

$$a\,x = a'(g + x).$$

Since x is very small compared with g, we can write

$$x = \frac{a'}{a}\,g.$$

In *fig. 1* the results are given in micrograms of Tedion-sulphur per leaf as a function of the time between the spraying of the tree and the taking of the leaf sample. Fig. 1*a* relates to the wettable powder, and fig. 1*b* to the miscible oil. It is seen that with both formulations the "external" sulphur, i.e. *on* the plant (△---△), drops in one to two days to half its initial value. (The fall-off in inherent activity of the ^{35}S — half-life 87 days — can be neglected here.) There was no rainfall after spraying until after the second leaf sample had been taken, so that this result cannot be attributed to rain. After about four days 20% of the powder preparation remains on the leaf and 40% of the miscible-oil preparation. Thereafter the amount of external sulphur diminishes slowly until, in about three months, there is virtually none left; the total activity is then equal to the

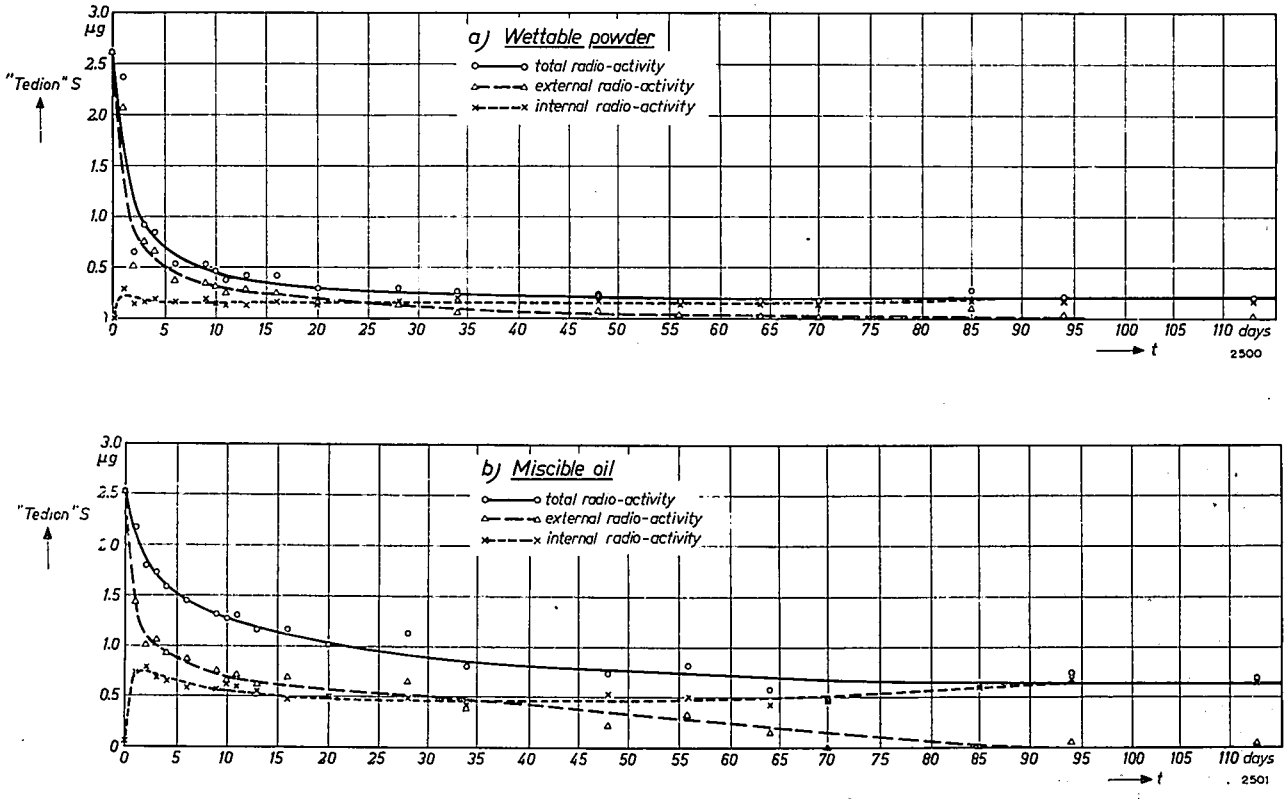


Fig. 1. Amount of Tedion-sulphur per leaf, as a function of the number of days following the spraying of apple trees with a preparation of *a*) wettable powder, and *b*) miscible oil.

"internal" activity, i.e. the activity of the sulphur inside the leaves.

The amount of this internal sulphur (× - - - ×) in fig. 1*a* and *b*) shows initially a rapid increase and in both formulations it reaches a maximum in about one day. This maximum is roughly equivalent to 8% of the activity of the total sprayed Tedion in the case of the wettable powder, and 30% in the case of the miscible oil. The leaves treated with the miscible-oil preparation thus take up almost four times as much Tedion as the leaves sprayed with the powder preparation (this proportion is also subject, of course, to other factors, such as the size of the particles on the leaf, the nature of the contact between the particles and leaf, etc.).

Beyond the maximum, the internal sulphur slowly diminishes to a value that remains constant for about a month; for the miscible-oil preparation this value is approximately three times higher than for the wettable powder. After about two months the internal sulphur showed a slight increase in both cases. This may be due to the fact that, at the end of the summer season, all life processes in the leaves — including those that break down the Tedion — begin to slow down, although the take-up of external residues presumably continues undiminished.

The method of isotopic dilution analysis, applied to the leaf samples treated with miscible-oil preparation, yielded the following results. In November 1957 (141 days after spraying) 79.2 micrograms of non-decomposed Tedion were found, with an activity of 0.157 microcuries ^{35}S. This is 17.5% of the total activity of the leaf sample, which, on the same date, was 0.9 microcuries.

The conclusions drawn from the leaf-sample experiments were the following. From the miscible-oil formulation three to four times as much Tedion is fixed on the leaf and taken-up inside the leaf as from the wettable-powder formulation. The miscible oil will thus be more economical in use. The Tedion penetrates quickly into the leaf. More than a month after treatment there is still a fair amount of non-decomposed Tedion left on and in the leaf, and even after two months enough remains to exercise an acaricidal action. The Tedion taken up in the leaf is subject to conversion and transport inside the plant, but it is continuously supplemented from the surface residue.

To evaluate these findings correctly, it must be realized that the eggs and larvae of mites are destroyed by contact with Tedion, and that the eggs laid by a fully-grown mite that has fed on the leaf sap will not hatch out if the Tedion content in the sap is sufficiently high. Effective mite control therefore requires the presence of Tedion *on* as well as *in* the leaf. Tedion is found to meet both requirements most satisfactorily.

Generally speaking, it is desirable that a pesticide applied to cultivated plants should be broken down by the plant at such a rate that little or none is left when the time comes for the plant or its fruits to be consumed by man or domestic animal. Tedion, which is in any case virtually innocuous to mammals, actually undergoes such a slow break-down.

We shall now touch on the experiments relating to the *fruits* of the sprayed trees. Apples, picked about 6½ weeks after spraying, were first washed in warm nitromethane to dissolve the surface Tedion. They were then peeled, after which the peelings and the fruits were separately reduced to pulp and the pulp masses extracted with a chloroform solvent. Only unchanged Tedion is extracted in this way. The two chloroform extracts were evaporated to dryness and the residues dissolved in nitromethane. Measurements were then made of the radioactivity of the three nitromethane solutions (but not of the extracted pulp). The results are given in *Table I.*

Table I. Measured radioactivity of apples 6½ weeks after spraying the trees with radioactive Tedion in different formulations.

Tedion formulation	Part of apples	Per five apples:	
		Microcuries of ^{35}S	Calculated Tedion content, μg
Miscible oil	On peel	0.127	30
	In peel	0.096	23
	In fruit (extract)	0.077	18
			71
Wettable powder	On peel	0.035	9
	In peel	0.013	3
	In fruit (extract)	0.033	8
			20

In all, five apples took up 71 μg of Tedion from the miscible oil and 20 μg from the wettable powder (i.e. roughly the same proportions as found in the leaves). Both amounts are too small to exercise an acaricidal action and *a fortiori* are insignificant as regards their possible toxicity to man.

Experiments on rats

The experiments on rats were done with a preparation consisting of a suspension of radioactive Tedion in an emulsion of salad oil in water. The specific radioactivity of the Tedion amounted to a few millicuries per gram. In the first experiments

the rats were given a total dose of 100, 50 or 10 milligrams of Tedion per kilogram of bodily weight, either at once or spread over several days. The preparation was introduced directly into the stomach through a catheter.

As a preliminary experiment a number of rats received daily one tenth of the total dose (100 mg) for ten days. Some of these rats were then killed, and some of their organs and tissues were examined for radioactivity by various chemical extraction techniques. The rats not killed were left under normal conditions in their cages and the activity of their faeces and urine was tested.

In a subsequent experiment a total amount of 100, 50 or 10 mg of radioactive Tedion per kg bodily weight was administered to rats in a single dose. After 48 hours these rats were killed and their organs and tissues, as well as their collective faeces and urine, were tested for radioactivity. Details of the methods employed will be described elsewhere.

The preliminary experiment showed that many tissues had become radioactive, especially fatty tissues, the tissue of the gastro-intestinal tract, muscles and liver. The hair, too, showed slight radioactivity, indicating that part of the Tedion had been broken down into simpler sulphur compounds from which, by biosynthesis, sulphur-amino acids had formed that are found in hair. By far the greater part of the Tedion, however, was found to have left the body of the rats: 71% with the faeces, 4% with the urine, and 7% was still present in the contents of the intestinal tract. It may therefore be assumed that 82% of the administered dose is excreted and that a mere 18% is taken up in various parts of the body. After the dosage was stopped the radioactivity in the living rats dropped rapidly to a negligible value. *Table II* gives a survey of the distribution found in a rat which had received 100 mg per kg in a single dose. The sum of the amounts of sulphur determined by radioactivity measurements (1236 μg) agrees remarkably well with the administered dose (1221 μg), a fact which strengthens our confidence in the method employed.

The same experiment was done on two other rats with a dose of 50 mg Tedion per kg, and on another two with 10 mg per kg. The relative distribution over the various constituents of the body differed very little from that shown in Table II.

Isotopic dilution analysis was used to ascertain what percentage of the measured radioactivity of the various body constituents was due to non-decomposed Tedion. Without going into details, we may mention the general conclusion that, of the administered 100 mg of Tedion per kg bodily weight, 40 to 45% was found unchanged after 48 hours, mainly in the faeces. Thus, 55 to 60% is broken down in one way or another, presumably to a large extent by the action of the bile.

Table II. Distribution of Tedion-sulphur in the body of a rat, 48 hours after receiving a dose of 13.6 mg of Tedion (i.e. 1221 μg of sulphur). The rat weighed 136 g.

Constituents analysed		Tedion-sulphur in μg
Faeces		878.8
Urine		51.3
Skin and hair	acetone extract (cutaneous fat)	87.8
	residue (mainly hair)	5.2
Gastro-intestinal tract tissue		84.4
Muscles		16.9
Fat		78.1
Blood (major part)		0.10
Brains		0.16
Heart		0.10
Kidneys		0.39
Liver		4.14
Lungs		0.43
Spleen		0.06
Carcass (unspecified remainder of body, largely bone, gristle and sinew)		24.3
Cage swill		4.23
	Total	1236.41 ± 20

The experiments described provide no answer to such questions as: Where in the plant or the body of the animal is the Tedion decomposed? and: What are the decomposition products? Investigations along these lines are in progress. In conclusion it may be remarked that the foregoing gives further evidence, if any were needed, of the great value of radioactive-tracer techniques to biological research.

Summary. Radioactive-tracer techniques are applied to investigations into the behaviour in plants and animals of "Tedion V 18" (an acaricide for the control of spider mites). The sulphur in the active ingredient of Tedion, which is 2,4,5,4′-tetrachlorodiphenyl sulphone, is partly replaced by the radioactive isotope ^{35}S. Apple trees were sprayed with preparations of this radioactive Tedion in two formulations: as a wettable powder and as a miscible oil. Radioactivity measurements showed that an active residue remains *on* the leaf for a long period, whilst the Tedion taken up *in* the leaf is subject to conversion and transport within the plant, being continuously supplemented from the surface residue. The miscible oil proves to be more economical in use than the wettable powder. Administered to rats (max. dose 100 mg per kg bodily weight), about 40 to 45% of the Tedion is found unchanged after 48 hours, mainly in the faeces; the remainder is broken down.

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 Laboratories, Eindhoven, The Netherlands.

2696: H. C. Hamaker: Attributenkeuring in theorie en praktijk (Statistica neerl. **13**, 37-58, 1959, No. 1). (Theory and practice of sampling by attributes; in Dutch.)

This paper brings a survey of present-day theory and practice of sampling by attributes. In section 2 the main points in the theory of OC-curves are recapitulated. Section 3 lists the various features which people have attempted to incorporate in sampling tables, and the various factors which influence our choice of a sampling plan. Subsequent sections discuss economic theories, the importance of the distribution of the percentage of defective items over the inspection lots, the relation between sample size and lot size, and some of the existing sampling tables. In a final section the possibilities for further improvements in existing sampling practices is considered. One conclusion is that in many situations a constant sample size regardless of lot size would have very definite advantages.

2697: A. Claassen and L. Bastings: The gravimetric determination of nickel with dimethylglyoxime in the presence of copper (Z. anal. Chemie **165**, 354-360, 1959, No. 5).

A detailed investigation has been made of the gravimetric determination of nickel with dimethylglyoxime in the presence of copper. Copper interference is eliminated by precipitation in a solution containing tartrate and thiosulphate at a *p*H of 5.5-6.5. The solubility of the nickel dimethylglyoxime complex has been determined as a function of temperature, *p*H and alcohol concentration.

R 385: K. Teer: Investigations into redundancy and possible bandwidth compression in television transmission (Philips Res. Repts. **14**, 501-556, 1959, No. 6).

In this thesis (Delft, Sept. 1959) the possibility of bandwidth compression in television transmission is considered. Three different aspects of redundancy present in normal television transmission are considered: (1) the statistical aspect, which is concerned with probability distributions of brightness, (2) the physiological aspect, which is concerned with the properties of the eye (resolving power, persistence of vision, differential sensitivity and colour perception), (3) the psychological aspect, which is related to levels of consciousness. After this analysis of redundancy, transmission systems of narrow bandwidth are described which have been investigated by the author. In all these systems bandwidth compression is effected by a decrease in the number of frames per second, viz., by decreasing the field frequency or the information per field. For practical realization of the former method a suitable memory device is needed. Considerations are restricted to the memory device, in particular to a vidicon-type camera tube, which can be used as a fairly simple device of this kind. A decrease of the information per field can be realized by use of dot-interlace and subcarrier techniques, which are examined in detail. Finally, the use of these principles in colour television is considered, mainly with the N.T.S.C. system and a two-subcarrier system.

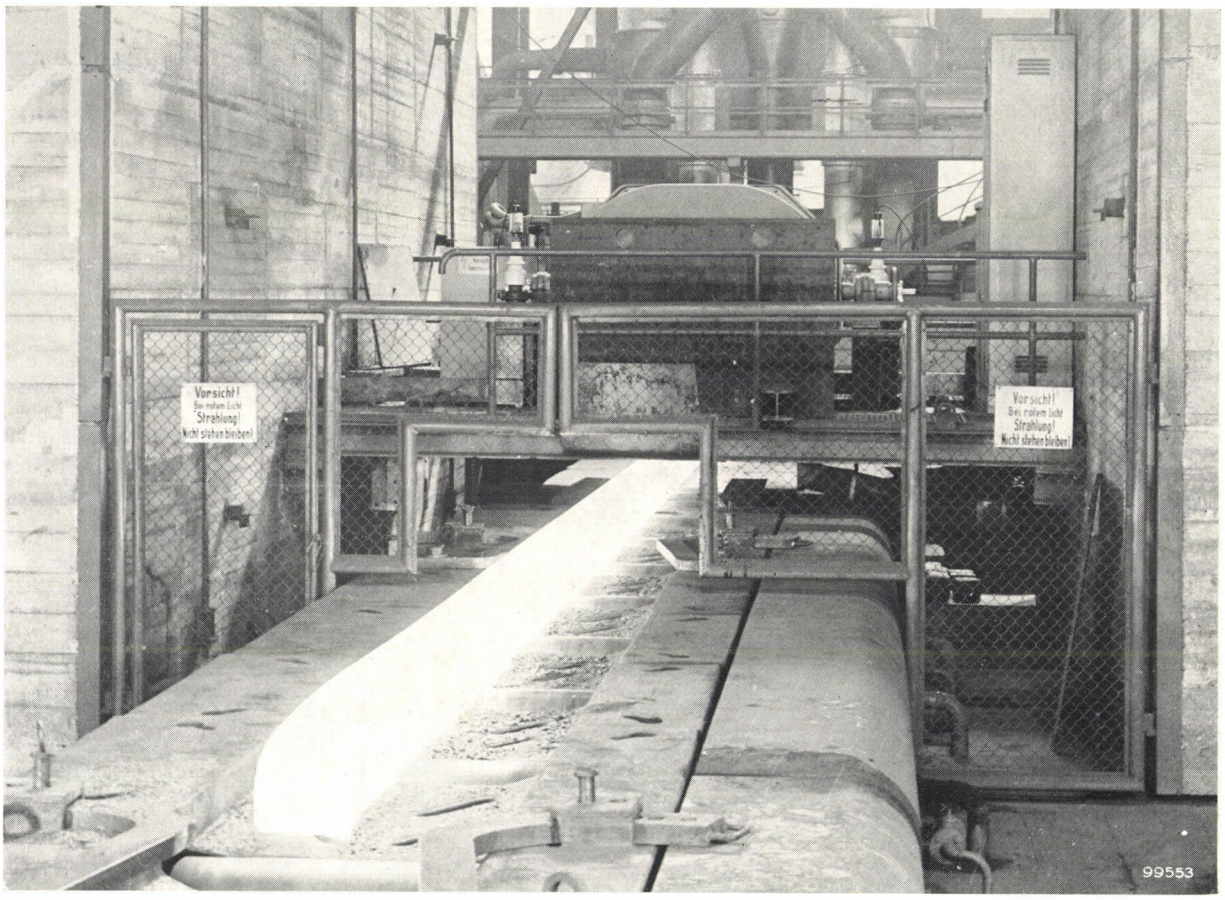
R 386: F. A. Kröger, F. H. Stieltjes and H. J. Vink: Thermodynamics and formulation of reactions involving imperfections in solids (Philips Res. Repts. **14**, 557-601, 1959, No. 6).

Two types of formulation of reactions involving imperfections in solids are analysed and compared with each other. As a result it appears that, in Schottky's system, the amount-of-substance variables mostly refer to relative building units. Therefore true thermodynamic potentials can be assigned to them and they also can appear as components in the phase rule. In an alternative system the amount-of-substance variables do not refer to building units, but to structure elements and to associates between them. To those no true thermodynamic potential can be assigned. However, by introducing virtual thermodynamic potentials for them, all problems can be dealt with, often in a much more convenient way than with the other type of formulation. Also, appearing as quasi-components in the phase rule, they greatly facilitate its handling. This is illustrated by treating with the aid of the structure-element system various cases of disorder. For the description of those properties of a crystal which are not related to the thermodynamics of reactions, the use of structure elements is also of advantage.

VOLUME 21, 1959/60, No. 10 pp. 281-316 Published 30th August 1960

Philips Technical Review

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



X-RAY INSPECTION OF HOT STEEL BILLETS DURING ROLLING

by W. J. OOSTERKAMP, J. PROPER and M. C. TEVES. 620.179.152: 621.771

In steelworks the molten steel prepared in open-hearth furnaces, Bessemer converters or electric furnaces is cast in iron moulds, in which it solidifies into ingots of several tons. After the ingots have set, but before further cooling, they are taken from the moulds and reduced in a roughing mill to billets or blooms about thirty feet long and $8'' \times 8''$ or more in cross-section. Frequently the billets are immediately subjected to further rolling operations to reduce them to bars or sections only an inch or so thick.

A problem encountered in this process is that blowholes are formed in the steel during solidification, which, if not plugged by the rolling, materially weaken the finished product. The same applies to slag inclusions in the steel. Experience has shown that defects of this kind are particularly prevalent in the front end (the "crop") of the billet, and it is therefore the invariable practice to cut off an end portion some metres in length and to discard it as scrap. Inspection of the cut face for holes or inclusions reveals whether too much or too little has been discarded. The presence of such defects, however, cannot be very clearly observed on the red-hot metal. Consequently it is quite possible that

in fact too little will be removed, which means that poor-quality steel is passed for further processing, or that, to be on the safe side, too much is cut off, which is a waste of good material.

It is evident that X-ray investigation of the hot billets in the rolling mill would result in a considerable saving of material and also make for a more reliable product. Furthermore, it would show whether defects were present in parts of the billet other than the crop, and finally — and this is perhaps the most important point — it would provide a practical and simple means of investigating the effect of various production factors on the formation of cavities and other casting defects. It was indeed similar considerations that led to the widespread adoption of X-radiography for the inspection of welds; apart from its great utility for routine inspections, radiography has also contributed substantially to the improvement of welding techniques [1]).

In the present case it is a matter of detecting holes or inclusions, measuring only a few millimetres across, in a steel billet at least 200 mm ($\sim 8''$) thick, and which, moreover, is moving at a steady rate of, say, a foot or two per second. For this purpose the normal X-ray equipment for industrial radiography, using X-radiation up to 300 keV, is hopelessly inadequate: the half-value layer of steel for X-ray energies of 300 keV is 6.5 mm, which means that this radiation is attenuated in steel 200 mm thick by a factor of about 2^{-30} or 10^{-9}!

At the steelworks of Phoenix-Rheinrohr AG, Düsseldorf, plans were developed to tackle this problem with the aid of various modern devices: a betatron as radiation source, an X-ray image intensifier as detector and an industrial closed-circuit television system for the final display. Preliminary tests and investigations, for which Philips made available a special X-ray image intensifier (designed for other purposes), led to the conclusion that the envisaged equipment would meet the requirements. One of the rolling plants at Phoenix-Rheinrohr was then equipped with a trial installation, and several thousand hot billets were subjected to X-ray fluoroscopy. Some particulars of this installation and of the results obtained are given below [2]).

The betatron radiation source, a product of Brown Boveri (Switzerland), generates X-ray energies up to 31 MeV. This extremely hard radiation can be used for fluoroscopy of relatively very thick materials, especially since the betatron gives an appreciable intensity, comparable to that of an X-ray tube working on 300 kV, 10 mA. What is more, an X-ray energy in the region of 30 million electronvolts is about the most favourable that can be chosen in this case, the penetrating power in steel then being greatest; at higher energies the absorption in steel progressively increases as a result of pair production. Nevertheless, the half-value layer of steel for 31 MeV radiation is only 34 mm, which is still much smaller than the object thickness. After traversing 200 mm of steel, the rays are reduced in intensity to about 2^{-6}, i.e. to roughly 1% of the initial value. Observation at this low residual intensity is made possible by the X-ray image intensifier, to which we shall refer presently. The third part of the installation is, as mentioned above, an industrial television system for remote display of the image formed on the viewing screen of the image-intensifier tube. This television equipment uses an image-orthicon camera tube and was supplied by Fernseh GmbH of Darmstadt. A full-size image of the viewing screen is formed in the usual way on the photocathode of the image orthicon by a conventional optical system. The use of television offers several advantages. In the first place it is the easiest answer to the problem of safely positioning the observer, who cannot of course be immediately behind the intensifier viewing screen, where the level of residual radiation is still dangerously high. Further, the television system, with its extremely sensitive image orthicon, enables a luminance gain to be obtained that supplements that of the image intensifier. Finally, it is a simple matter in a television circuit to provide for *contrast intensification* [3]). All these favourable factors make it possible in the installation described to achieve better detail perceptibility than if the viewing screen were observed by means of mirrors and a telescope — a method which, although safe for the observer, is hardly a reasonable proposition in a steel-rolling plant.

Since a uniform brightness is obtained in the whole field of view during the screening of the rectangular-sectioned steel billets, a considerable degree of contrast intensification may be applied. To obtain a uniform intensity distribution across the X-ray beam from the betatron, a suitably shaped equalizing absorber is placed in front of the betatron window.

---

[1]) An alternative and widely used method of non-destructive testing is by means of ultrasonic vibrations; this is not suitable for examining hot billets, however, because of excessive damping in the material at the temperatures involved.

[2]) For a detailed report see W. Lückerath, K. Fink and R. Flossmann, Stahl und Eisen **79**, 1637, 1959 (No. 22).

[3]) Contrast control, but usually with the object of attenuating instead of intensifying contrast, is common practice in television, a gamma corrector circuit being used for the purpose. See e.g. Philips tech. Rev. **15**, 227 *et seq.*, 1953/54. The circuit there employed permits a contrast gain of up to 1.5×. Other circuits are used for obtaining the much higher contrast intensification required in the present case.

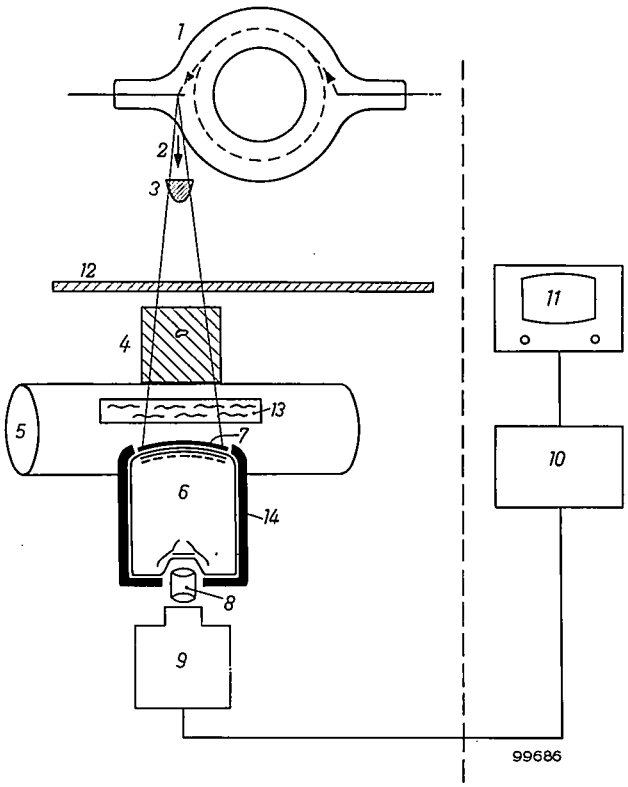


Fig. 1. Schematic cross-section of the trial installation at Phoenix-Rheinrohr, Düsseldorf, for visual X-ray inspection of hot steel billets during rolling. *1* betatron tube. *2* beam of 31 MeV X-rays. *3* copper equalizing absorber. *4* steel billet, which moves along a line of rollers *5* under the betatron. *6* X-ray image intensifier tube with lead layer *7* over aluminium window. *8* optical system. *9* television camera tube. *10* television video circuits. *11* display tube. *12* aluminium plate which shields the betatron from the heat radiated by the hot steel. *13* water-cooled heat-radiation shield for the image-intensifier tube. *14* lead shield to prevent scattered X-rays from penetrating into the image intensifier.

The layout of the installation is illustrated in *figs. 1* and *2*, and some additional details are mentioned in the captions.

The X-ray image intensifier works on the principle described in previous articles in this journal. X-rays falling on a fluorescent screen excite an image, the photons of which release electrons from a photocathode in contact with this fluorescent screen. These electrons are accelerated by a high potential (25 kV) and focused on a second fluorescent screen, the viewing screen. The image intensifier placed at the disposal of Phoenix-Rheinrohr (see *fig. 3*) had a tube wall made entirely of aluminium and gave an electron-optical reduction of more than 10 times, and an overall luminance gain of about 1200 times. The field of the image intensifier, i.e. the diameter of the X-ray fluorescent screen, is 9″; this is the same as that of the recently introduced image intensifier for medical applications. It is thus an intermediate size between the types earlier described, of 5″ and 11″ [4]). For use in conjunction with the betatron, a layer of lead 2 mm thick was fitted over the aluminium window of the intensifier. This was done

[4]) M. C. Teves and T. Tol, Electronic intensification of fluoroscopic images, Philips tech. Rev. **14**, 33-43, 1952/53 (description of 5″ tube). The 11″ tube is described in Philips tech. Rev. **20**, 331-345, 1958/59 (No. 11).

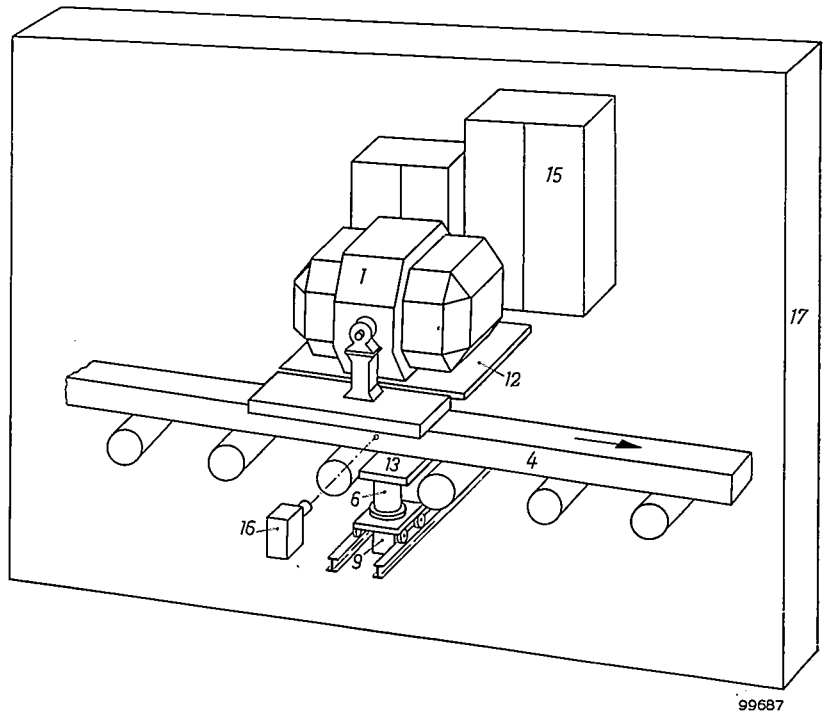


Fig. 2. Perspective sketch of the installation. Meaning of numbers as in fig. 1. Other numbers: *15* power pack for betatron. *16* second television camera, enabling observer to keep an eye on the movement of the billet. *17* one of the concrete walls that help to shield the environment against X-radiation scattered from the billet (see also title photograph). The image intensifier *6* and the camera tube *9* are mounted on a carriage to permit lateral displacement.

because only a very small fraction of the 31 MeV radiation is directly absorbed by the X-ray screen and used for producing the fluorescent light. In the lead, the 31 MeV X-rays release electrons that possess an energy of about 10 MeV and travel in directions not very different from the beam of X-rays.

The extremely high gain obtained in the X-ray image intensifier (and in the television circuit) does not alter the fact that the information, upon arrival in the image intensifier, is carried by a relatively low number of X-ray quanta. A high "noise" level in the picture is therefore inevitable [5]). The effect is

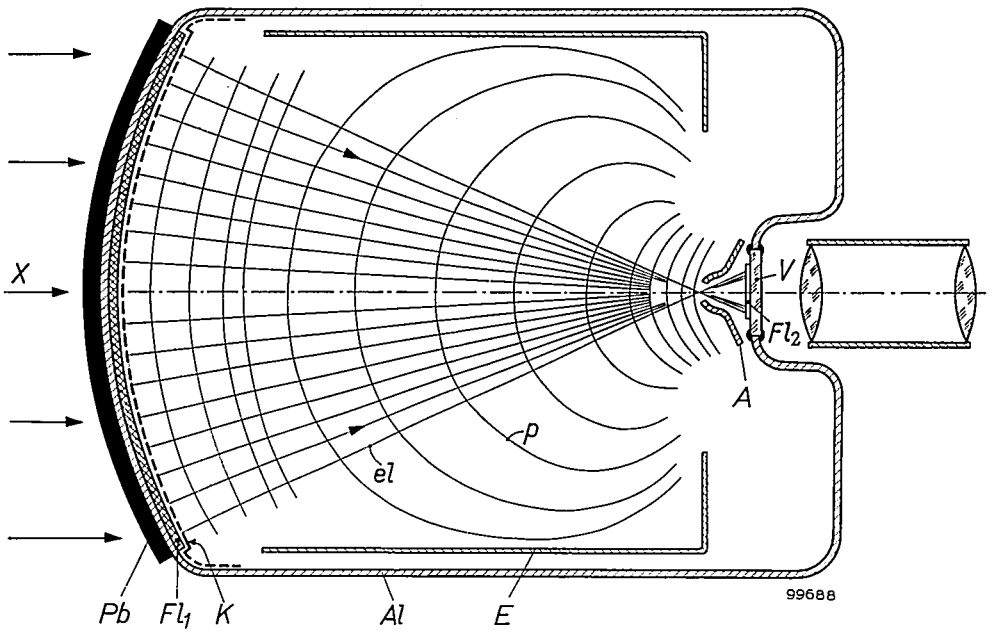


Fig. 3. Cross-section of the Philips 9″ X-ray image intensifier placed at the disposal of Phoenix-Rheinrohr AG. *Al* aluminium wall of tube. Fl_1 X-ray screen. *K* photocathode. *E* focusing electrode. *A* anode. Fl_2 viewing screen. *V* glass viewing window. A number of electron paths (*el*) and the cross-sections of a number of equipotential surfaces (*p*) are shown. For the purpose of the experiment a lead layer *Pb* was fitted over the aluminium wall covering the X-ray screen: the incident 31 MeV X-radiation (*X*) releases electrons from this layer.

These electrons pass through the aluminium window without appreciable scattering (the window is only 1 mm thick) and in the fluorescent screen, fitted on the inside of the window, they excite about 4000 photons of fluorescent light per electron. Image intensification then occurs as already outlined above, and each 10 MeV electron from the lead gives rise finally to 400 000 light photons from the viewing screen.

It is worth pausing here to reflect on the remarkable sequence of particle interactions occurring in this equipment. *Electrons*, accelerated in the betatron, generate *X-ray photons* in the target, and these, after passing through the object, liberate in their turn *electrons* from the lead shield. These strike the X-ray screen, where they excite *light photons*, which again release *electrons* from the photocathode. These electrons, after acceleration, finally produce *light photons* again in the viewing screen. In the television system there is the further sequence of electronic phenomena by which the picture information is transmitted to the picture tube, where it is again converted to light photons that finally reach the eye of the observer.

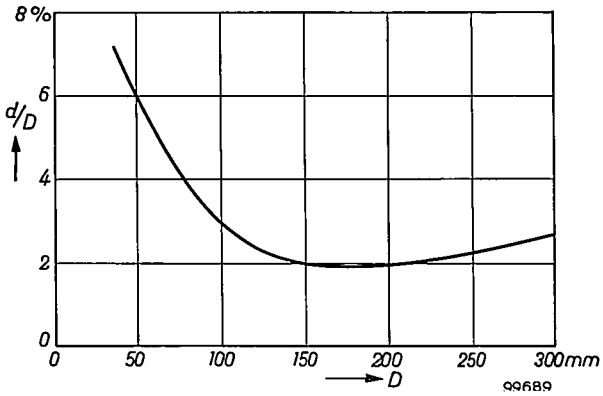


Fig. 4. Perceptibility of thin wires on a steel billet (penetrameter test) during inspection with the installation described. The percentage thickness (*d*/*D*) of the thinnest wire still just perceptible is plotted as a function of the thickness *D* of the steel.

[5]) For a more detailed consideration of noise in X-ray images see: T. Tol and W. J. Oosterkamp, The application of the X-ray image intensifier, II. The perception of small object-detail, Philips tech. Rev. 17, 71-77, 1955/56.

reduced to some extent by the afterglow of the television picture tube, in which each image element is integrated over a certain time. Of course, the afterglow persistence may not be unduly long in view of the fact that the steel billet is a moving object.

In spite of the remaining noise, the trial installation gave surprisingly good results. Penetrameter measurements (i.e. determination of the thinnest wires still just perceptible) yielded the curve shown in *fig. 4*. This indicates that, in a fluoroscopic examination of a steel billet 200 mm (8″) thick, steel wires of 4 mm diameter on the billet can still be observed. A statistical analysis was made of the casting defects found during the screening of several thousand hot steel billets at Phoenix-Rheinrohr, against the defects found during the subsequent inspection of cold sawn-off sections. The conclusion was that screening in this way discloses 50% of all blowholes 3 mm thick and 90% of all blowholes 7 mm thick.

Summary. In steelworks the quality of the finished product can be improved and needless waste of material avoided if the presence of blowholes and inclusions can be detected *during* the rolling of the hot billets (cross-section of the order of 8″×8″). To this end a trial installation was designed and constructed by Phoenix-Rheinrohr AG, Düsseldorf, in which the steel billets were moved through a beam of 31 MeV X-rays generated in a betatron, and the image observed with the aid of a special X-ray image intensifier (field of 9″ diameter) and an industrial television system. The results of screening several thousand hot billets showed that 50% of all holes 3 mm thick where detected, and 90% of all holes 7 mm thick.

NUCLEAR MAGNETIC RESONANCE

by D. J. KROON.

539.143.42.082.722.56:620.18

The phenomenon of nuclear magnetic resonance in solids and liquids yields absorption spectra from which valuable information on their structures can be obtained. The article below first gives a simple exposition of the theory, and then describes in more detail the equipment required and a few applications. The latter illustrate the usefulness of the method both for pure scientific research and for routine investigations in various industries.

Introduction

Amongst the most recent tools in the investigation of condensed phases (solids and liquids) are the methods of measurement that depend on the absorption of electromagnetic energy at specific wavelengths in the metre and centimetre regions. According to the mechanism causing the absorption, the methods can be divided into three groups: paramagnetic resonance, nuclear magnetic resonance and cyclotron resonance.

In paramagnetic and nuclear magnetic resonance the absorption is caused by the presence in the substance of elementary magnets, i.e. the "spins" of electrons and atomic nuclei which, in the presence of a constant external magnetic field, are subjected to an *alternating magnetic* field. Cyclotron resonance occurs when more or less free electrons (e.g. in semiconductors), in the presence of a constant external magnetic field, are subjected to an alternating *electric* field. In this article we shall be concerned only with nuclear magnetic resonance. The related effect of paramagnetic resonance was described in an article published some time ago in this journal [1]), to which we shall refer when occasion arises. An appendix to the present article deals in some detail with the points of difference between paramagnetic and nuclear magnetic resonance.

As mentioned, nuclear magnetic resonance occurs when atomic nuclei exhibit a spin, that is to say, they have a net angular momentum and magnetic moment. This is so with many kinds of nuclei, in principle in all cases where the number of neutrons and the number of protons in the nucleus are not both even. Familiar examples of nuclei that show a readily measurable effect, i.e. a marked absorption, are ^{19}F, ^{23}Na, ^{27}Al and ^{31}P, but most especially ^{1}H, the hydrogen nucleus or proton.

The information that nuclear magnetic resonance can provide is of various kinds. Hydrogen atoms are present in many inorganic solids (e.g. in molecules of water of crystallization) and in nearly all organic molecules. The form in which the hydrogen nuclei are present has an influence on the resonance spectrum, i.e. the spectral distribution of the measured absorption. In an organic liquid, for example, each kind of molecule containing hydrogen shows a characteristic line spectrum, which can therefore serve for identifying the molecule. In this way it is possible to analyse mixtures of organic compounds. Nuclear magnetic resonance equipment is for this reason regularly used for routine analyses in the petroleum and oil-refining industries and the plastics industry. In solids the resonance frequency of the H nucleus is very broad. From the spectral distribution of the absorption here, one can draw conclusions regarding the ordering of H ions and other ions in the crystal and also regarding the motion of electrons and ions, diffusion phenomena, and so on.

An earlier article in this Review explained how nuclear magnetic resonance can also be used for the accurate measurement of magnetic fields [2]). We shall touch on this application when describing the equipment.

In the following pages we shall first give a simple theory of the phenomenon of nuclear magnetic resonance. After describing the measuring equipment we shall consider, by way of illustration, some examples of problems that have been studied by the method of nuclear magnetic resonance. When dealing with the equipment we shall distinguish between measurements on solids and on liquids. Owing to the different nature of their spectra, these two states of aggregation call for different techniques of measurement.

For the determination of crystal structures the well-established methods of X-ray diffraction and infra-red absorption can also be used. It is important to note, however,

[1]) J. S. van Wieringen, Paramagnetic resonance, Philips tech. Rev. **19**, 301-313, 1957/58.

[2]) H. G. Beljers, Three methods of measuring magnetic fields, III. Measurement by the proton-resonance method, Philips tech. Rev. **15**, 55-62, 1953/54.

that these methods are not so suitable when it is a matter of detecting and localizing hydrogen atoms.

X-ray diffraction depends on the scattering of X-rays by the atoms or ions of which a crystal lattice is built up. This scattering decreases rapidly the lower the atomic number of the atom or ion. Scattering by H ions is so weak that information on these ions can be obtained from X-ray-diffraction measurements only by indirect means. The related method of neutron diffraction is better in this respect, and in fact has proved very useful [3]). It can only be adopted, however, in laboratories associated with a suitable nuclear reactor. Moreover, in these investigations it is often desirable to replace the hydrogen ions (protons) by ions of heavy hydrogen (deuterons), which of course is only practicable in special cases [4]).

The infra-red absorption of hydrogen atoms bound to oxygen (OH vibrations) is readily measurable. The frequency of these vibrations depends markedly on the surroundings of the OH group in the lattice, and it is frequently possible to deduce the configurations in which the OH group occurs in the crystal. The interpretation of the measurements, however, is very complicated and the information obtained is generally insufficient for a complete determination of the structure. Furthermore, the sample under measurement is required to be transparent to some extent, which often calls for the use of special techniques.

The method of nuclear magnetic resonance is thus a useful complement to the older methods of structure determination

Historical note

As long ago as 1936 the Dutch physicist C. J Gorter predicted the possibility that electromagnetic waves would be absorbed in a solid through the agency of the atomic nuclei if these were subjected to an external magnetic field. His experiments at the time produced no results. It later appeared that the substances on which the experiments were done happened to show only a very small effect, and this was the main reason for the failure of the experiment. Some years later, thanks to the advances in high-frequency techniques, the physicists F. Bloch and E. M. Purcell, working independently and with different methods of measurement, were able to confirm Gorter's theory. In 1952 both these American workers were awarded the Nobel prize for physics.

Concise formulation of the theory of nuclear magnetic resonance

In general an atomic nucleus possesses, apart from mass and electric charge, an angular momentum and a magnetic moment. In the classical model, the atomic nucleus is imagined to behave like a rapidly spinning magnetic top, the magnetization being in the direction of the axis of rotation. Quantum-mechanical considerations require that if the nucleus be subjected to a constant external magnetic field, the axis of rotation of the top, which precesses about the field direction, can assume only certain discrete orientations with respect to the direction of the magnetic field. The number of possible orientations depends on the magnitude of the angular momentum. For hydrogen nuclei (protons) there are two possible orientations, viz. in the direction of the field and opposite thereto. Other nuclei have two or more possible orientations (fluorine 2, sodium 4, bismuth 10).

The energy of the nuclear magnet in the magnetic field is:

$$E = \mu B \cos \vartheta,$$

where μ is the magnitude of the magnetic moment, B the field strength (induction) and ϑ the angle between the elementary magnet and the external field (*fig. 1*). The possible values of $\cos \vartheta$ are given by the formula:

$$\cos \vartheta = \frac{m_I}{\sqrt{I(I+1)}},$$

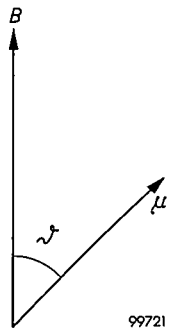


Fig. 1. Magnetic dipole (moment μ) in a magnetic field B.

where I, the spin quantum number of the nucleus, is a multiple of $\frac{1}{2}$, and m_I can assume a number $(2I+1)$ of discrete values, each differing by 1, the extremes being I and $-I$. The energy E can also be written in the form:

$$E = \mu_z B \frac{m_I}{I},$$

where $\mu_z = \mu\sqrt{I/(I+1)}$ represents the largest possible projection of μ on the axis of the field. The magnetic moment μ is given by:

$$\mu = g\sqrt{I(I+1)}\,\frac{e\,h}{4\,\pi m_p}, \quad \ldots \quad (1)$$

where e and m_p represent respectively the charge and mass of the proton, and g is a numerical constant

[3]) G. E. Bacon, Neutron diffraction, Oxford Univ. Press, London 1955. See also G. E. Bacon, Ned. T. Natuurk. **25**, 5, 1959 (No. 1).

[4]) G. R. Ringo, Neutron diffraction and interference, Section 36, Handb. Physik, Vol. **32**, Springer, Berlin 1957, p. 603. J. A. Goedkoop and B. O. Loopstra, Ned. T. Natuurk. **25**, 29, 1959 (No. 2).

that differs according to the type of nucleus (in general g is between 0.1 and 10).

The spin quantum number for hydrogen nuclei is $I = \frac{1}{2}$. There are thus, as remarked, only two states possible. The energy difference between these states is

$$\Delta E = 2\mu_z B. \quad \ldots\ldots \quad (2)$$

In an external field $B = 0.5$ Wb/m^2 [5000 gauss] this energy difference is of the order of 10^{-26} J (10^{-7} electronvolt).

If the nucleus is subjected simultaneously to a constant magnetic field and an alternating magnetic field at right angles thereto, there is a possibility that the atomic nucleus will change from the one state to the other, whereby an amount of energy equal to the difference ΔE will be absorbed from the alternating field or given up to it. The condition for the occurrence of such transitions is that the frequency f of the alternating field should satisfy:

$$f = \frac{\Delta E}{h}, \quad \ldots\ldots\ldots \quad (3)$$

where h is Planck's constant ($h = 6.62 \times 10^{-34}$ Js).

Eliminating ΔE from the expressions (2) and (3) gives the condition for resonance absorption as

$$f = \frac{2\mu_z}{h} B, \quad \ldots\ldots\ldots \quad (4a)$$

or, after substitution of μ_z (from (1)) and putting $I = \frac{1}{2}$:

$$f = g \frac{e}{4\pi m_p} B. \quad \ldots\ldots \quad (4b)$$

For hydrogen nuclei μ_z is 1.4×10^{-26} A.m^2. In a field of $B = 0.5$ Wb/m^2, as mentioned above, we thus have $f = 21.3$ Mc/s.

The procedure for observing the absorption is basically as follows. A sample of the substance having sufficient nuclei of the type under investigation is introduced into a coil which is connected to a weak high-frequency source of roughly the resonance frequency f_0. The coil is placed between the poles of a magnet and the magnetic field is slowly varied. When the magnetic field B_0 reaches the value given by (4a) — for the frequency f_0 — resonance occurs, giving rise to a slight energy absorption which results in a change in the Q of the coil. By including the latter in a suitable circuit, e.g. a bridge circuit, the change in the Q can be measured and recorded as a function of B, to give the magnetic absorption spectrum. The effect of relaxation phenomena on the absorption will be touched on presently.

Structure of the absorption spectrum

From the mechanism described it might be inferred that resonance absorption occurs only at a single frequency, determined by (4a) or, if the frequency f_0 is given, at a certain induction B_0, given by:

$$B_0 = \frac{h}{2\mu_z} f_0.$$

The experiment shows, however, that absorption occurs in a certain region of B values around B_0, and that the recorded absorption line thus has a certain width, and sometimes a certain structure. This may partly be attributed to the fact that, whereas a field B has been applied and measured, the nuclei investigated may in reality be in a field that differs from B by a small amount b. The resonance condition is then:

$$B + b = \frac{h}{2\mu_z} f_0 = B_0,$$

or

$$B = B_0 - b.$$

Absorption thus occurs at a value B of the measuring field that differs from B_0. If b differs for various parts of the specimen or for various kinds of nucleus, the result is a broadened line or a line spectrum. One may also conceive of performing the experiment such that the external field (B_0) is kept constant and the frequency f varied. In that case resonance would occur when $f = f_0 + \Delta f = (2\mu_z/h)(B_0 + b)$, so that $\Delta f = (2\mu_z/h)b$. The value $\Delta f = 1$ c/s corresponds to $b = 2 \times 10^{-8}$ Wb/m^2.

Various reasons can be given for the occurrence of the field-strength deviations b. The first is quite trivial, being of an instrumental nature and connected with the fact that the magnetic field is not perfectly uniform. As a result, parts of the specimen are found in a field that is not exactly identical with the field B. This source of broadening can be reduced to negligible proportions by using small specimens and a magnet whose air gap is narrow compared with the diameter of the pole pieces. This applies at least to solids in which, for other reasons (to be discussed), marked broadening effects occur in any case. In liquids, which generally exhibit a spectrum having sharp resonance lines, the instrumental broadening may not be negligible.

Another cause of instrumental broadening is found especially when recording the spectra of solids. As we shall see, in such cases the measurement is performed by finding an average value over a range δB of field strengths. Since δB cannot be

made infinitely small, this also gives rise to line broadening (see later, fig. 10*a*).

The other causes of line broadening and of the occurrence of line spectra, discussed in the following, are essentially concerned with the physical conditions to which the nuclei are subject during the experiment, and about which the absorption spectrum can therefore provide information.

Local fields in crystals

For the present we shall confine the discussion to solids, that is to crystals. In a crystal the field acting on the nuclei consists of the external field B (now assumed to be uniform) together with the dipole fields of neighbouring nuclei. For hydrogen nuclei this dipole field at a distance of 1 Å is of the order of 10^{-3} Wb/m^2. The total field acting on a nucleus subjected to a dipole field b is $B + b$. At resonance we again have $B + b = B_0$, hence:

$$B = B_0 - b.$$

The resonance line is then (in terms of B) displaced over a distance b.

A case frequently encountered is that of two neighbouring protons. This is found when investigating the absorption of two protons of a molecule of water of crystallization (H-O-H). The field b caused by the one proton at the position of the other depends on the distance r between the two protons and on the angle ϑ which the line connecting them makes with the external field B. The dipole field is given by:

$$b = \pm \frac{\mu_z}{r^3}(3\cos^2\vartheta - 1). \quad . \; . \; . \quad (5)$$

The $\pm$ sign arises because the "perturbing" proton can be oriented in two ways in the field B. For one pair of protons we then obtain, instead of a single resonance peak, two peaks at a distance $2b$ apart. We shall see presently that, as regards the magnitude of the effect, this statement requires some modification.

This, then, is the situation when we have a single crystal and consider the proton pairs therein which take up a specific orientation with respect to the crystal axes. If we turn from a single crystal to a powder, we find that the angle ϑ for the kind of proton pairs considered has a different value in each crystallite of the powder. The resonance pattern is consequently smeared out to a broad band in which, however, two peaks can still be recognized.

So far our discussion has taken account only of neighbouring protons forming a particular H-H pair, for example the protons of a single H-O-H molecule. These protons will also, however, be subjected to the dipole fields of protons of other H-O-H molecules and in addition to the dipole fields of other crystal nuclei. This causes a further broadening of the absorption pattern. As an example *fig. 2* shows the proton-resonance absorption of a powdered specimen of potassium fluoride ($KF.2H_2O$).

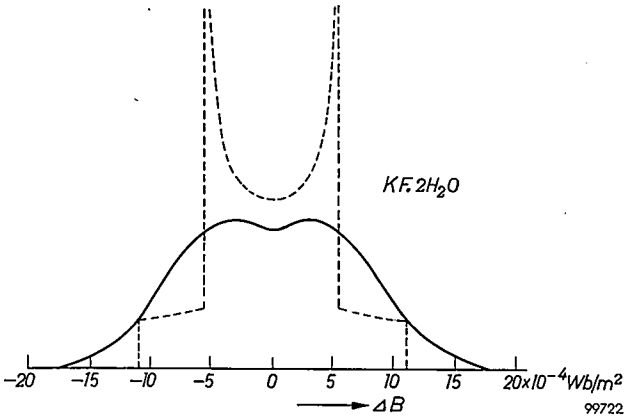


Fig. 2. Spectrum of the absorption caused by proton resonance in a powder sample of potassium fluoride ($KF.2H_2O$).
The dashed line represents the theoretical shape of the spectrum if there were only interaction between the protons of a single molecule of water. For calculating this curve the distance between the peaks was taken as 11×10^{-4} Wb/m^2, and the area below the curve was made equal to that under the experimental curve.

In the simple case of isolated proton pairs it is a fairly simple matter to predict a theoretical absorption curve that compares reasonably with the experimental result. In more complicated cases this is not so easy. In such cases, only the "second moment" of the broadened line is determined. If the line shape is given by

$$y = \mathrm{f}(B),$$

its second moment is given by

$$\overline{(\Delta B)^2} = \int (B - B_0)^2 y \, \mathrm{d}B \Big/ \int y \, \mathrm{d}B.$$

If the value of $\overline{(\Delta B)^2}$ has been found experimentally, it can be compared with the theoretical value, for which Van Vleck derived the following [5]):

$$\overline{(\Delta B)^2} = C \sum_k \sum_{j \neq k} (3\cos^2\vartheta_{jk} - 1)^2 \, r_{jk}^{-6}.$$

Here r_{jk} represents the distance of a perturbing proton (j) to the perturbed proton (k), ϑ_{jk} the angle which the vector r makes with the external field, and C a constant comprising, among other things, the magnetic moments of the nuclei. The summation should really be extended over all the protons of the

[5]) J. H. van Vleck, Phys. Rev. 74, 1168, 1948.

crystal, but in practice it is sufficient to take only a small volume element of about 10 Å diameter.

Let us now return for a moment to the magnitude of the broadening. Apart from the "static" mutual interaction of the ions, referred to above, there is also a dynamic interaction, resulting from the fact that the protons, owing to their precession about the direction of the field, generate local alternating fields as well as a local static field. It would be going too far to deal with the complete quantum-mechanical treatment of this problem, which takes into account both the static and the dynamic effects [6]). The result, however, is fairly simple, and states that the expression $(3\cos^2\vartheta - 1)$ occurring in the formula must be multiplied by a factor of $\frac{3}{2}$, making the total broadening $\frac{3}{2}$ times the static broadening.

If we consider only two protons, we see that they can orient themselves in three ways with respect to the field: both parallel, one parallel and the other anti-parallel, and both anti-parallel. This gives rise to three energy levels each being $\Delta E = 2\mu_z B$ apart (see formula 2).

The mutual interaction described above (analogous to two coupled pendulums) perturbs the energy levels. As a result, the separation of the energy levels is now:

$$\Delta E = 2\mu_z B_0 \pm 3\mu_z^2 r^{-3}(3\cos^2\vartheta - 1).$$

This means that resonance is found for

$$B = B_0 \pm \tfrac{3}{2}\mu_z r^{-3}(3\cos^2\vartheta - 1), \quad \ldots\ldots \quad (6)$$

which, except for the factor $\frac{3}{2}$, is the same as (5). The factor $\frac{3}{2}$ does not occur in the broadening caused by the influence of other nuclei on the protons, for which the static calculation gives the correct result.

Interaction with the crystal lattice

Apart from the interaction between the protons mutually and between the protons and other nuclei, interactions occur between the protons and the atoms of the specimen as a whole. This too gives rise to additional line broadening.

This additional broadening can be explained with the aid of Heisenberg's uncertainty principle. This tells us that the energy ΔE of an excited state is indeterminate by an amount δE, given by:

$$\delta E\,\delta t \approx h. \quad \ldots\ldots\ldots \quad (7)$$

The quantity δt in the present case is a "relaxation time" τ, the significance of which is as follows. In the absence of an external field the spin directions are randomly distributed owing to thermal equilibrium. If, at a given moment, all spins were made parallel, then as a result of interaction with the lattice this ordering would relax and ultimately vanish according to an exponential law:

$$e^{-t/\tau}.$$

In the case of protons in a liquid τ is of the order of 10 sec and therefore $\tau^{-1} = 0.1$ c/s. With the aid of (7) we find that this corresponds to an additional broadening of approximately 2×10^{-9} Wb/m^2. In solids τ may have a wide range of values, but the additional broadening here is almost invariably negligible compared with the other causes of line broadening.

Thus, although the interaction with the lattice or with the environment is often negligible as regards the effect on the line width, it is really essential to the occurrence of resonance absorption. In an external field B in which there are N protons, the number of parallel-aligned protons (N_1, state 1) at thermal equilibrium is greater than the number of anti-parallel-aligned protons (N_2, state 2), to an extent given by:

$$\frac{N_2}{N_1} = \exp(-\Delta E/kT), \quad \ldots\ldots \quad (8)$$

where T is the temperature and k is Boltzmann's constant ($k = 1.38\times10^{-23}$ J/°K). If, now, an alternating field of the desired frequency is applied, and if there were no interaction with the lattice, after some time a state would arise where $N_1 = N_2 = \frac{1}{2}N$. From that moment onwards as many protons would go over per unit time from state 1 to state 2 as from state 2 to state 1, and no absorption would be observed. The interaction with the lattice, however, tries to restore the situation (8); consequently, even when an alternating field is present, N_1 remains greater than N_2, and absorption can, therefore, still be observed. Also, it follows from this that, for the purpose of detecting the absorption, the alternating field should not be too strong. These points were already discussed in connection with paramagnetic resonance in the article quoted under [1]).

Structure of the spectrum in liquids

In contrast to crystals, where the line widths may cover 10^{-4} to 10^{-2} Wb/m^2, the resonance lines found for liquids are very narrow, of the order of, e.g., 10^{-7} Wb/m^2. Special measures, to be discussed later, are therefore necessary to keep the instrumental width (non-uniform field) below this limit. The fact that the marked line broadening found in crystals does not occur in liquids is due to Brownian motion which, in classical terms, so frequently perturbs the phase of the precessional movement that only the effect of the average field remains, the action of

[6]) See, for example, E. R. Andrew, Nuclear magnetic resonance, Cambridge Univ. Press, London 1955, pp. 152-154, 240-242.

neighbouring fields being eliminated by averaging with respect to time. The interaction with the medium remains as the sole cause of line broadening. As in crystals, however, this is a very slight effect.

The result of this situation is that, at field strengths of, say, 10^{-5} Wb/m^2, displacements and splittings of a resonance line are observed which are due to effects that may also be present in the solid but escape observation because of the large total broadening in the case of solids. These effects are largely attributable to the action of the magnetic fields of electrons. It is due to them that the spectra of various organic compounds differ so markedly from one another and can even be used for identifying these compounds. We shall deal with this subject at greater length when discussing the practical examples.

Equipment for nuclear-magnetic-resonance investigations

Investigations on liquids

As remarked, for observing nuclear magnetic absorption in liquids the magnetic field should be highly uniform. The field variation over the dimensions of the specimen (a few millimetres) should be less than $1 : 10^6$.

To obtain a sufficiently uniform magnetic field for these kinds of experiment, magnets are used, as mentioned above, in which the air gap is narrow compared to the diameter of the poles of the magnet. The pole pieces should be made of highly homogeneous material and ground optically flat. The remaining non-uniformity of the field is corrected by passing a current through suitably-positioned flat coils on the pole faces.

With the object of achieving a still higher resolution, the specimen is often rapidly rotated. This has a narrowing effect simular to that discussed above as regards the influence of the Brownian motion on the internal fields: the differences in macroscopic field strength are eliminated by averaging with respect to time. In this way a resolution of 10^{-8} Wb/m^2 can be achieved.

During the measurement the magnetic field should remain constant. In the case of an electromagnet, a highly stabilized current should be used, constant to within 10^{-6} of the desired value. Residual, relatively small fluctuations due to other sources are compensated by means of a coil of many turns wound round the magnet poles. The e.m.f. induced in this coil by a fluctuating magnetic field is amplified and converted into a current which is fed in anti-phase to a correction coil on the magnet. In this way a constancy of 3 in 10^8 can be maintained for several minutes.

An easier method to attain constant magnetic fields is to use a permanent magnet. The only important cause of deviations of the magnetic field in this case is that the magnetization of the magnetic material (e.g. "Ticonal") is temperature-dependent. The temperature of the magnet should change by no more than 10^{-4} °C if a constancy of 1 in 10^8 is to be attained. The magnet is therefore placed in a thermostat. Owing to the high thermal capacity of the magnet, a simple thermostat is sufficient to cope with the fluctuations of the ambient temperature.

The frequency of the applied alternating field, too, should have a stability of 1 in 10^8. This can be obtained by using a crystal-controlled oscillator placed in a thermostat. If a number of such crystals are available, various kinds of nuclei can be investigated. Since, however, as mentioned, nuclear magnetic resonance is mainly used for the investigation of organic compounds, the resonance of protons is of primary interest.

The variation of the magnetic field required for scanning the spectrum is produced by means of an additional excitation coil, through which a current is passed that rises very slowly and linearly from zero to a maximum (sawtooth waveform). In the case of a liquid spectrometer the maximum value of the field variation need only be quite small (a few 10^{-4} Wb/m^2).

Various types of high-resolution spectrometers are described in the literature, and some are also commercially available. As an example we shall mention some particulars of the high-resolution spectrometer marketed by Mullard (type SL 44 Mo2). This uses a permanent magnet which is accurately maintained at a constant temperature. The magnet gives a field of 0.9 Wb/m^2 and has poles of 15 cm diameter and a gap 3 cm in length. The resonance detector consists of a double-T measuring bridge, which is connected to a high-frequency amplifier, followed by a diode detector. The output voltage of this detector, which is proportional to the output voltage of the bridge, is applied to an oscilloscope having a long-persistence screen and to a recording unit. The bridge is fed by a crystal-controlled oscillator of 40 Mc/s. The specimen holder consists of a glass tube of 3 mm diameter, which is filled with the substance for investigation and rapidly rotated. The photograph in *fig. 3* gives an impression of the equipment. A block diagram appears in *fig. 4*; some additional details are given in the caption.

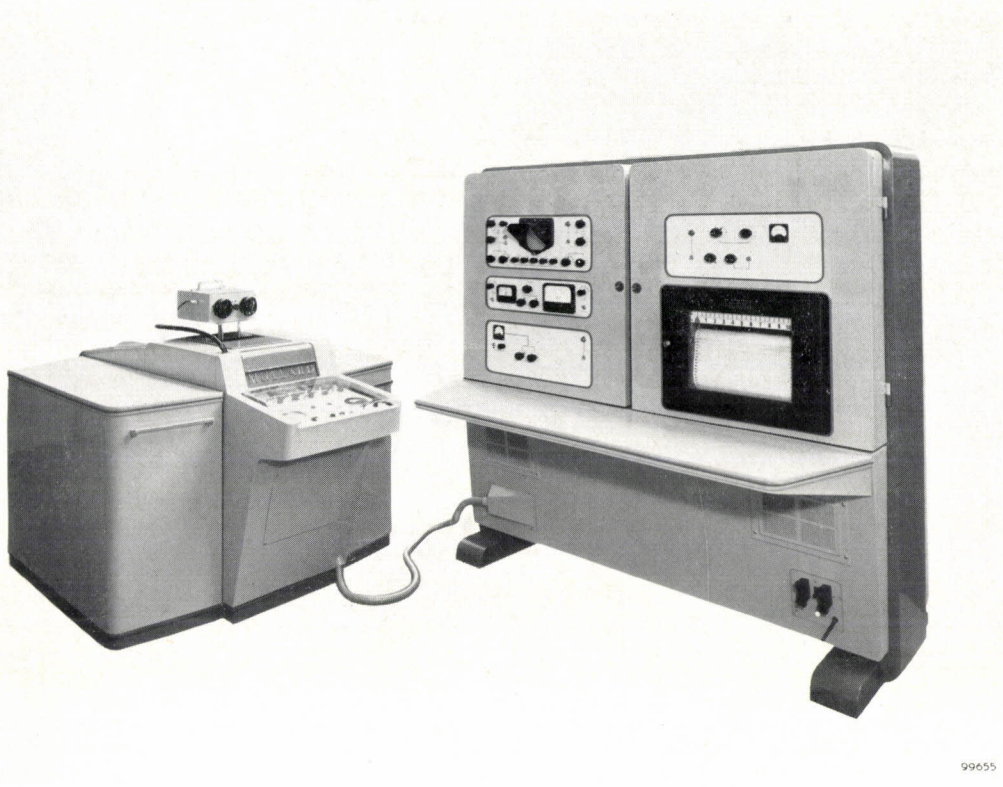


Fig. 3. Mullard equipment for nuclear-magnetic-resonance investigations on liquids (high-resolution spectrometer, type SL 44 Mo2).

Investigations on solids

Since the width of the lines in solids is generally very large (10^{-4} to 10^{-2} Wb/m^2), a highly uniform field is not necessary for absorption measurements on solids, nor need the field be so highly constant with time or the measuring frequency so stable.

On the other hand the absorption at a given field strength is now roughly a hundred times weaker. Compared with liquids, therefore, larger specimens are used and more sensitive equipment. The higher sensitivity is achieved by a different measuring procedure (see below). Higher demands are also made on the versatility of the spectrometer, since in solid-state investigations it may be necessary to measure the resonance of a large number of elements that may occur in crystals. Whilst the study of protons is often important for the determination of crystal structures (water of crystallization), the resonance of other nuclei (Cu, Al, etc.) enters into the study of *metals*. Investigation of the structure of the resonance line as a function of temperature, for example, makes possible the study of a number of effects (such as atom movements, which make the resonance line narrower, and diffusion phenomena).

As an example of a solid-state spectrometer we shall discuss the equipment used in the Philips Research Laboratories at Eindhoven. A block diagram of the set-up is shown in *fig. 5*.

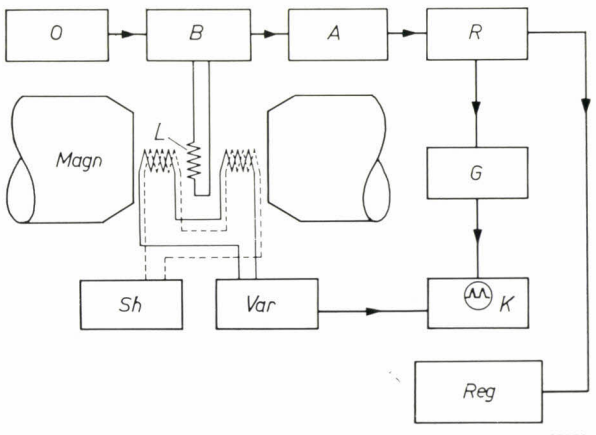


Fig. 4. Block diagram of the Mullard spectrometer in fig. 3 used for liquid analysis. *Magn* permanent magnet (induction in air gap 0.9 Wb/m^2). *L* alternating-field coil in which the sample is placed. *O* crystal oscillator, frequency 40 Mc/s, for generating the high-frequency alternating magnetic field inside the coil *L*. *B* measuring bridge. *A* preamplifier. *R* receiver. *G* diode detector. *K* cathode-ray oscilloscope. *Reg* recorder. *Var* generator of auxiliary current for the sawtooth variation of the static magnetic field used for scanning the spectrum. The current source *Sh* is used for correcting the magnetic field in the air gap to match it to the oscillating frequency of the crystal.

The field for this spectrometer is again generated by a permanent magnet. The fairly considerable range of field strengths required for the analysis of solids is obtained here by means of gaps of variable length in the yoke. With the aid of a cam the yoke bars can be raised, thereby decreasing the magnetic field. A particular advantage of this construction is that, within a certain range, the field can be reproducibly adjusted to any desired value, so that no auxiliary field is needed for matching to the oscillator frequency (cf. *Sh* in fig. 4).

The cam is driven by a synchronous motor via a gearbox, enabling the spectrum to be continuously scanned. The shape of the cams is such as to ensure the most linear possible variation of the field with time. As in the case of a high-resolution spectrometer, this is desirable in order that equal distances on the uniformly moving strip chart in the recorder will correspond to equal variations of B.

To the slow field variation a low-frequency sinusoidal modulation of very small amplitude is added (coils L_{mod} and oscillator O_{mod} in fig. 5). *Fig. 6* shows what happens. A portion of the resonance line of width δB is scanned. The resultant signal, when δB is sufficiently small, is proportional to the derivative dy/dB of the absorption curve $y = f(B)$.

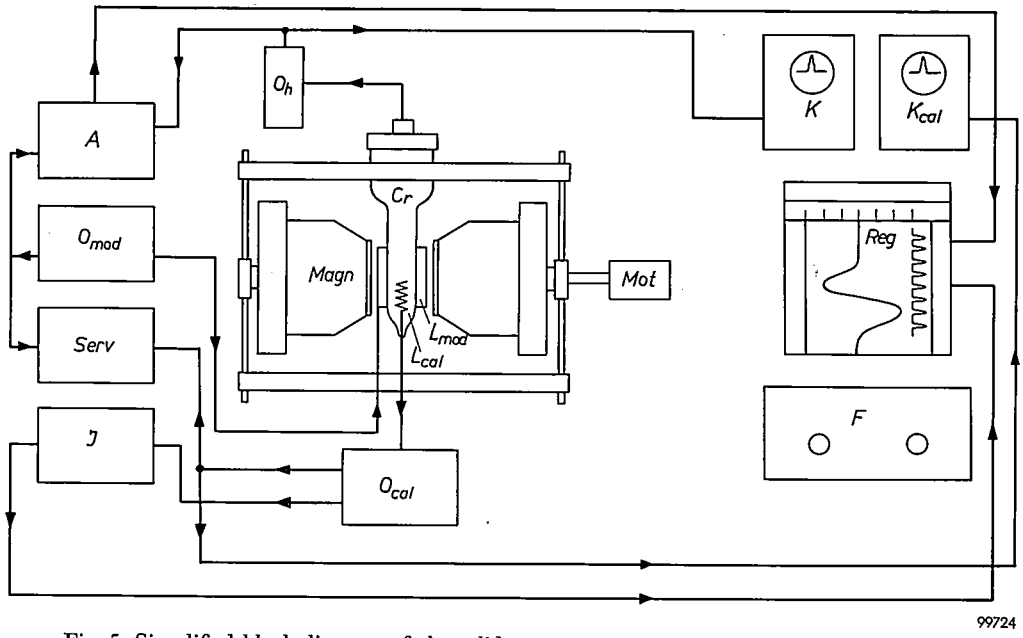


Fig. 5. Simplified block diagram of the solid-state spectrometer in Eindhoven. *Magn* permanent magnet (induction in air gap 0.45 Wb/m²). *Mot* synchronous motor with gearbox for shifting the yoke bars of the magnet, thereby effecting a continuous change of the field in the air gap (spectrum scanning). *Cr* cryostat, containing alternating-field coil and sample. O_h oscillator for generating the high-frequency alternating magnetic field in this coil. L_{mod} coils for the low-frequency modulation of the static magnetic field, fed by the oscillator O_{mod} (230 c/s). *A* selective amplifier. *K* cathode-ray oscilloscope for observing the output signal of O_h. *Reg* recorder. L_{cal} calibrating coil, used in conjunction with oscillator O_{cal} for calibrating the magnetic field by means of proton resonance. *Serv* servo amplifier, which keeps the calibrating oscillator tuned to proton resonance; this can be checked on the oscilloscope K_{cal}. *J* crystal-controlled oscillator with mixer stage, supplying the calibration pulses to the recording apparatus. *F* frequency meter.

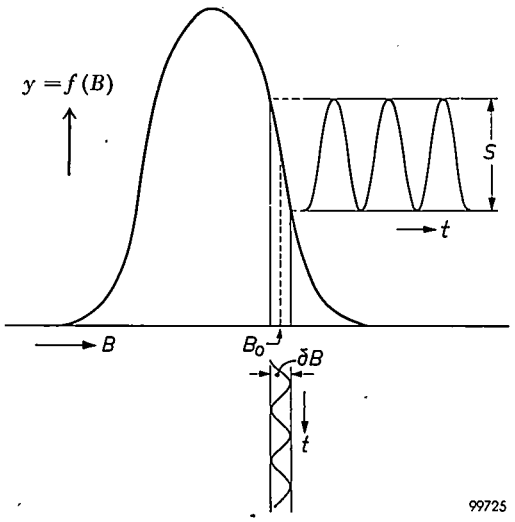


Fig. 6. Illustrating the scanning of a broad line by a low-frequency alternating magnetic field. The result, the derivative dy/dB of the absorption $y = f(B)$, is recorded. δB = modulation width, S = resultant signal.

This method of measurement has the advantage over the simpler method used for liquids that amplification is now possible in a very narrow band around the modulation frequency (230 c/s). As a result the unavoidable noise from the amplifier is suppressed and a considerable gain is achieved in the signal-to-noise ratio.

This method cannot be used with the liquid spectrometer because the line width, expressed in terms of frequency, is so very small (0.1 c/s). Modulation with a practical frequency, which must be at least thirty or forty c/s, would give rise to undesirable line broadening and satellite lines. However, since the absorption intensity in liquids is so much greater, a lower sensitivity is sufficient in their case.

The signal is detected, in the present equipment, with the aid of the same oscillator [7]) that supplies the high-frequency voltage to the alternating-field coil. This is done by making the coil form part of the frequency-determining network of the oscillator. A small change in the properties of the coil (the Q changes depending on the absorption of the specimen inside it) causes a considerable change in the voltage of the oscillator.

After detection the low-frequency signal is amplified in a narrow-band amplifier (A in fig. 5) and mixed with an alternating voltage whose frequency is equal to the modulation frequency. This finally produces a D.C. voltage which is proportional to the derivative of the resonance peak and is recorded on a strip chart. Examples of such recordings appear in fig. 10.

A second specimen and coil are present in the magnetic field. By means of proton resonance in this (liquid) specimen the magnetic field is continuously measured and calibrated by comparison with the signal from a crystal-controlled oscillator (J in fig. 5). The calibration marks appear on the recording. In the present equipment the distance between the marks is 1.17×10^{-4} Wb/m^2. This again clearly demonstrates the difference between the spectra of liquids and solids. Broadly speaking, about 20 calibration marks appear on one spectral line in the case of solids, whereas there may be 20 complete liquid spectra between two such calibration marks.

For measurements at low temperatures the complete specimen holder can be placed in a cryostat. With the set-up now in use, measurements can be made at temperatures as low as 20 °K; the recent availability of liquid He now makes it possible to measure down to 1.2 °K.

[7]) H. A. Thomas, Electronics **25**, January 1952, p. 114.

The caption to fig. 5 gives further particulars of the equipment. The photograph in *fig. 7* gives an impression of the set-up. The permanent magnet is shown separately in *fig. 8*, showing the mechanism for varying the air gap.

In the adjoining table some properties of the two nuclear magnetic spectrometers discussed are set out side by side.

	High-resolution spectrometer	Solid-state spectrometer
Specimen size	10 mm^3	1000 mm^3
Non-uniformity of the magnetic field over the whole of the specimen	$1:10^6$	$1:5\times10^4$
Instability of the static magnetic field with time	$1:10^8$	$1:10^5$
Instability of the alternating field	$1:10^8$	$1:10^5$
Resolving power $\Delta B/B$	$1:2\times10^7$	$1:5\times10^4$
Suitable for the nuclei:	especially ^{1}H, ^{19}F, ^{31}P	numerous
Modulation of static magnetic field	sawtooth	sawtooth + low-freq. sine wave
Amplitude of sawtooth	0.0002 Wb/m^2	0.03 Wb/m^2
Recorded quantities	absorption $y = f(B)$	dy/dB
Measurement at low temperatures	no	yes

There now follow some simple examples of the practical application of nuclear magnetic resonance.

Examples

The spectrum of ethyl alcohol

The ethyl-alcohol spectrum reproduced in *fig. 9* was recorded with a Mullard spectrometer. The spectrum consists of a number of very fine lines, and its total width is only 5×10^{-6} Wb/m^2. Since it is the spectrum of a liquid, the marked broadening usually found in solid spectra is lacking. The lines are due to hydrogen nuclei in the sample.

Let us now examine the structure of the spectrum more closely. There are three groups of lines. The integrated areas of each of the three groups are found to be in the ratio 3 : 2 : 1, corresponding to the formula CH_3-CH_2-OH. Why have these three groups a slightly different frequency and why is each group split into components (viz. 3, 4 and 1)?

The magnetic field at the position of one of the protons of the molecule is composed of the externally applied field plus the field produced by the surround-

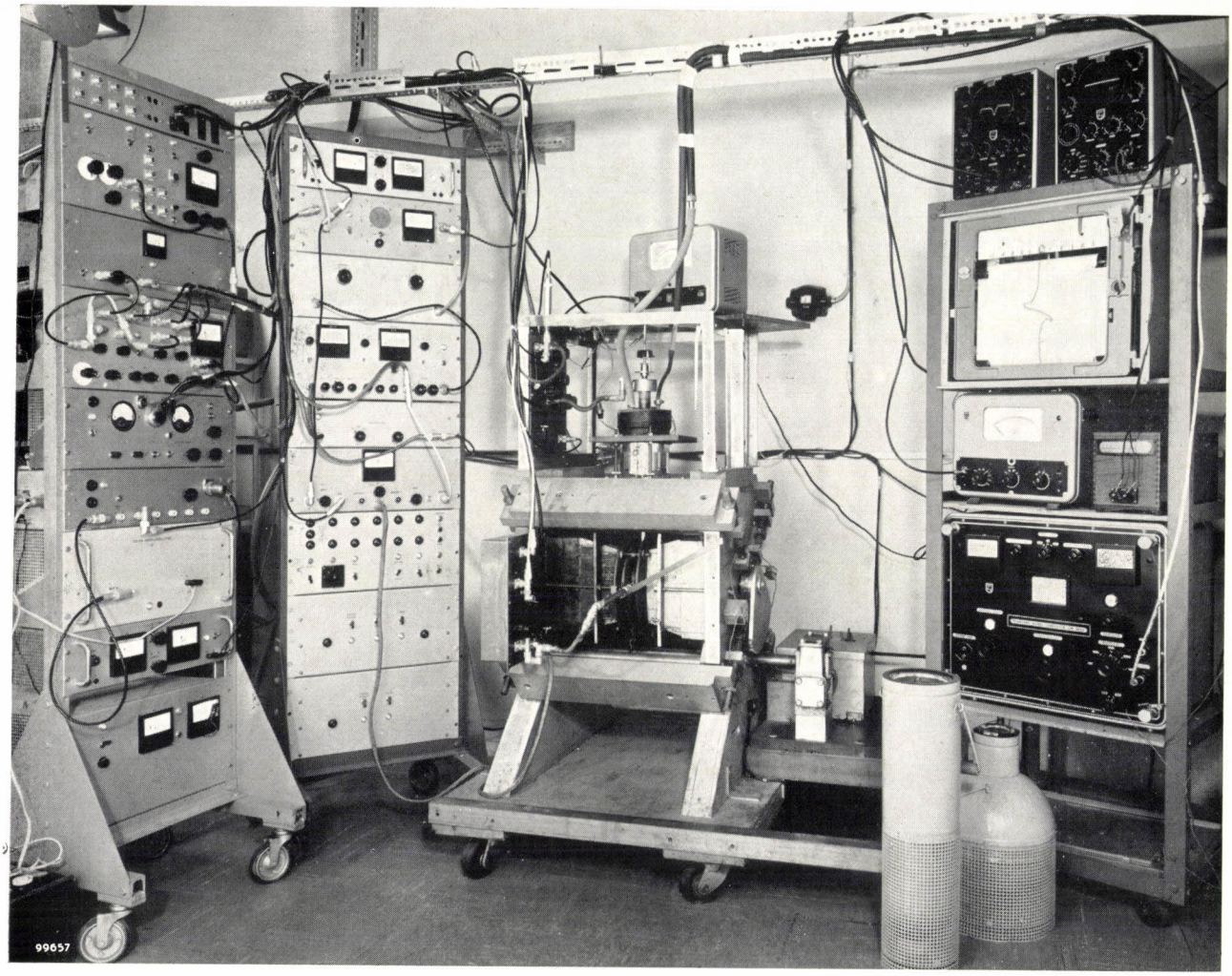


Fig. 7. Apparatus for the study of nuclear magnetic resonance in solids, in use at the Philips Research Laboratories, Eindhoven.

ing *valency electrons*. The electron density in the molecule is lowest in the OH group and highest in the CH_3 group, and for this reason the electron field differs for the three groups that make up the molecule. This causes the splitting into three groups of lines.

The splitting into components of a single group of lines arises as follows. The electrons move through the entire molecule and, as it were, communicate information to the nuclei regarding their immediate environment. Since, for example, the protons of the CH_2 group can assume different orientations in relation to the external field, the electron orbits can be polarized in different ways by the spin field of these protons. This gives rise to a change, albeit a very small one, in the magnetic field which the electrons produce at the position of the neighbouring groups. The two protons of the CH_2 group may either be both parallel with the measuring field, or one may be parallel and the other anti-parallel, or both anti-parallel. The total magnetic moment of

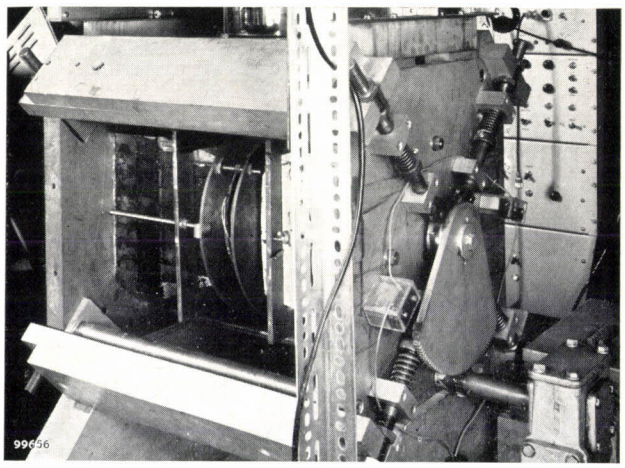


Fig. 8. The permanent magnet in the solid-state spectrometer at Eindhoven. On the right-hand side can be seen the toothed segment which is slowly turned by the motor via a gearbox (extreme right); attached to the segment is a cam which pushes the four radially mounted pins outwards. These raise slightly the four yoke bars, two of which can be seen in the photograph, thereby lengthening the total air gap in the magnetic circuit. This produces the desired time-linear variation of the magnetic field in which the sample is situated.

Fig. 9. Spectrum of ethyl alcohol (CH_3-CH_2-OH), recorded with the Mullard spectrometer (fig. 3). In the spectrometer used, the magnetic field was 0.75 Wb/m^2.

the CH_2 group is then respectively $2\mu_z$, zero or $-2\mu_z$. The probabilities of these three cases occurring are in the ratio of 1 : 2 : 1. This causes the line belonging to the protons of the CH_3 group to be split into three components with an intensity ratio of 1 : 2 : 1. In the same way we can argue that the line pertaining to the protons of the CH_2 group is split, under the influence of the neighbouring CH_3, into four components with an intensity ratio of 1 : 3 : 3 : 1.

Special considerations apply to the line due to the OH group. One might expect that this line, under the influence of the neighbouring CH_2 group, would also be split into three components of intensity ratio 1 : 2 : 1. Only one line, however, is in fact found. The explanation is that the alcohol sample used contains *impurities* in the form of traces of water, acid or base. Since the electron density around the OH group is small, the proton of the OH group can easily break away. Where traces of impurity are present there is therefore a constant exchange between the H ion of the OH group and the H ions of the impurity. This proton exchange occurs for every molecule very many times per second. The time during which a proton of the OH group is under the influence of the polarized electron orbits in the alcohol molecule is therefore extremely short, and the average frequency of the exchange is high compared with the frequency difference between the splitting components of the line group, which would exist if there were no such exchange (approx. 10 c/s). As a result there is no effective splitting into components. In extremely pure alcohol the line of the OH group does show the expected splitting into three components.

The further details visible in the spectrum are due to second-order effects which do not enter into the scope of this simple explanation and will not be discussed.

It should be noted that the described spin-spin interaction, via the polarization of the electron orbits, differs essentially from the direct spin-spin interaction which occurs in solids and which, in our case, is made ineffective by the Brownian motion of the molecules.

The above inferences drawn from the nuclear magnetic resonance spectrum are an elegant confirmation of the structure of the alcohol molecule, already determined by other means. It is evident that in other cases valuable information can be obtained from liquid spectra, that can serve for determining an unknown structure.

The other examples to be discussed refer to solids. They concern investigations made, or still in progress, in Philips Research Laboratories at Eindhoven, and in all cases involve proton resonance.

Hardening of "Araldite"

The hardening of "Araldite" (Ciba, Basle) is a striking example of the transition of a fluid into the solid state. Araldite is a widely used synthetic adhesive of the epoxy-resin type. The adhesive consists of two fluid components, the epoxy resin and a hardening agent added and mixed just before use. After about 30 hours the adhesive has set hard as a result of polymerization of the epoxy molecules. Some spectra, recorded at various times during the hardening process, are shown in *figs. 10a, b* and *c*. It can be seen how the original very narrow liquid

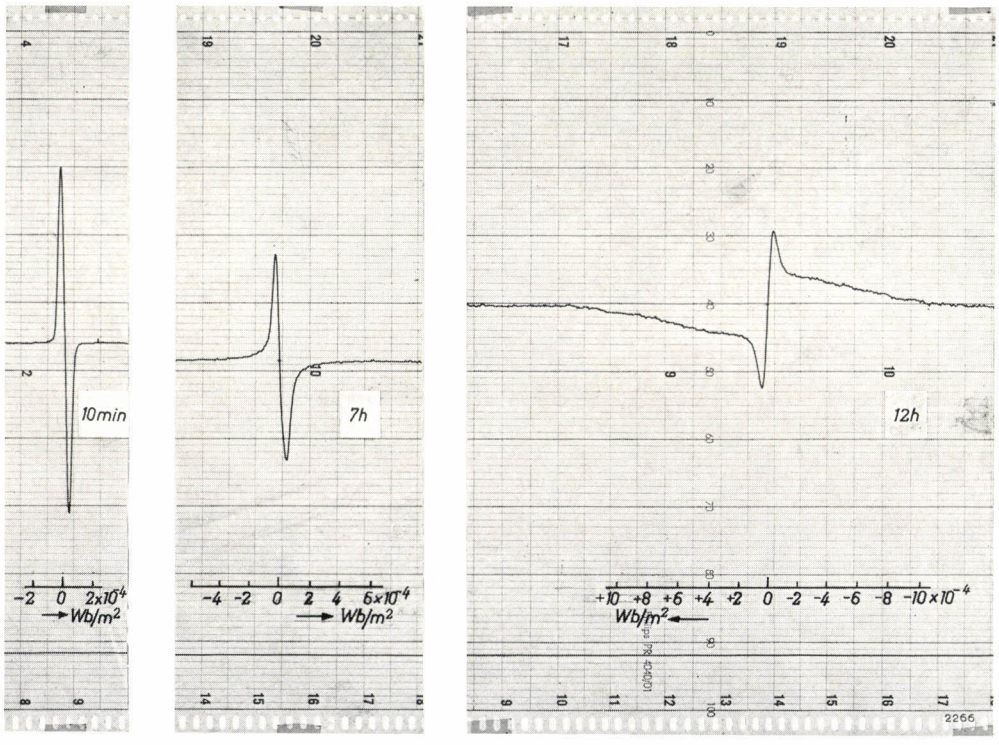
line, whose width is determined by instrumental causes, develops into a solid-state line having a width of about 2×10^{-3} Wb/m^2. The variation of the second moment with time indicates that the line broadening takes place in several more or less distinct stages. This allows inferences to be drawn regarding the gradual slowing-down of the motions of various groups of atoms in the molecule. This subject has been studied more extensively by Lösche [8]).

Motion of hydrogen gas in metals [9])

The following example refers to the motion of hydrogen gas in metals. Some metals and alloys can absorb relatively large amounts of gas, e.g. hydrogen, and for this reason they can be used as getters in radio valves. One of these alloys is thorium-aluminium, Th_2Al. The crystal structure of this alloy has been extensively studied [10]), and it is found that a maximum of four atoms of hydrogen is taken up per molecule of Th_2Al. Between the metal atoms are two interstices, each of which can entrap two H atoms. We shall consider the compound Th_2AlH_2, in which half the number of available interstices are filled.

At room temperature a very narrow resonance line was found with this compound. Partly from the intensity distribution in the line it was concluded that this narrow line is due to protons that jump very rapidly from one interstice to a neighbouring unoccupied one. Owing to this movement the dipole-dipole interaction between neighbouring protons is averaged out. At low temperatures ($T < 100$ °K) a line width of 8×10^{-4} Wb/m^2 was found. Since 8×10^{-4} Wb/m^2 corresponds to a frequency difference of about 4×10^4 c/s, it may be inferred from this that under these circumstances the time spent by a proton in an interstice is long in relation to 3×10^{-5} sec. Measurement of the line width as a function of T produced the curve shown in *fig. 11*. From the temperature at which the line-narrowing begins, in conjunction with the value of the line width at low temperature, it is easy to determine the way in which the average jump frequency ν_s

[8]) A. Lösche, Arch. Sciences **10**, fasc. spéc., p. 197, 1957.
[9]) D. J. Kroon, C. van de Stolpe and J. H. N. van Vucht, Arch. Sciences **12**, fasc. spéc., p. 156, 1959.
[10]) P. B. Braun and J. H. N. van Vucht, Acta cryst. **8**, 246, 1955.



a *b* *c*

Fig. 10. Recordings of the derivative of the absorption of "Araldite" during hardening. *a*) Ten minutes after mixing. The width of the line is determined by the modulation width δB. *b*) Seven hours after mixing. The foot of the curve already shows some broadening. *c*) Twelve hours after mixing. The curve shows marked broadening.
The recordings were made with arbitrarily decreasing or increasing magnetic field; in (*a*) and (*b*) the field happened to be increasing, in (*c*) decreasing.

depends on the temperature. With a minor simplification this can be described by the formula:

$$\nu_s \propto \exp(-E_a/kT),$$

where E_a represents the activation energy of the diffusion process. This is found to be equal to 0.22 eV. From the intensity distribution of the absorption line it can be inferred that at low temperature short-range ordering of the protons exists. This investigation is being pursued.

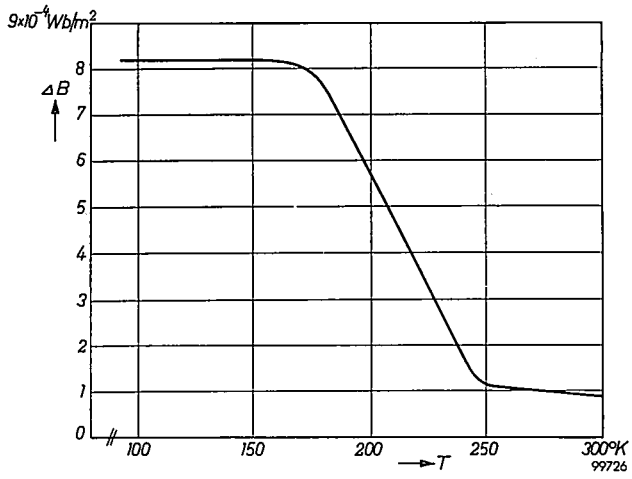


Fig. 11. Line width of the resonance in Th_2AlH_2 as a function of temperature.

Positions of protons in aluminium hydroxides [11])

To illustrate the application of the second-moment determination discussed above, we shall consider an investigation of aluminium hydroxide. Various hydroxides of Al are known, each with its own crystal structure; examples are hydrargillite (γ-$Al(OH)_3$), bayerite (α-$Al(OH)_3$) and boehmite (γ-AlOOH). We shall deal here with the first compound mentioned. The O atoms in this compound form a triangular layer structure. There are several conceivable sites for the H atoms; two are represented in *fig. 12a* and *b*. The theoretical value of the second moment (with an O-H distance of 0.96 Å) is 26×10^{-8} $(Wb/m^2)^2$ in the case *a*, and 29.4×10^{-8} $(Wb/m^2)^2$ in the case *b*. The value found by experiment is $(26.5 \pm 1.2)\times10^{-8}$ $(Wb/m^2)^2$, which argues in favour of the first assumption. Other crystal models that have been proposed give a widely different value, e.g. 15×10^{-8} $(Wb/m^2)^2$, and must therefore be rejected.

Finally, it may be noted that the few, very simplified examples of applications discussed here are by no means representative of all the possibilities offered by nuclear magnetic resonance. A list of the publications that have appeared so far on the subject of nuclear magnetic resonance would take up just as many pages as this article.

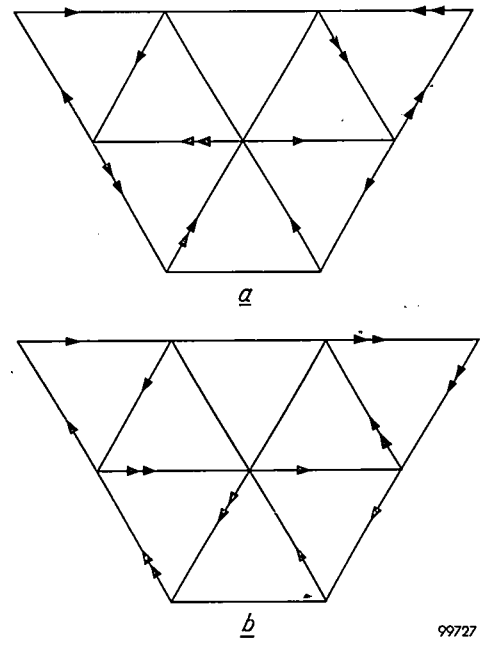


Fig. 12. In a crystal of γ-$Al(OH)_3$ layers of Al atoms alternate with double layers of O atoms. In the figure the O atoms are located at the corners of the triangles: in each case there are two O atoms one above the other at a distance apart of 2.79 Å. The H atoms are found between the two O layers; there are several theoretical possibilities as regards their precise position. (*a*) and (*b*) represent two possible configurations. In both cases each H atom lies on the oblique line connecting an O atom in one layer to a neighbouring O atom in the other layer, at a distance of 0.96 Å from one of these O atoms (the O-H bond length). These O-H bonds are indicated by the tail-ends of the arrows: a single arrow represents a bond with an O atom in the lower layer, and a double arrow a bond with an O atom in the upper layer. The result of nuclear magnetic absorption analysis argues in favour of the configuration (*a*).

Appendix: Comparison of nuclear magnetic resonance and paramagnetic resonance

In the introduction to this article mention was made of the relation between the phenomena of nuclear magnetic resonance and paramagnetic resonance. It is perhaps useful to touch here on some of the essential points of difference, some of which have already appeared in the foregoing pages.

The measurement techniques used for studying the two effects differ primarily in the wavelengths employed: paramagnetic-resonance investigations are performed at centimetric wavelengths, and thus use klystrons, waveguides and so on. Nuclear magnetic resonance is studied at wavelengths of some tens of metres, which calls for ordinary radio valves and transmission lines. This frequency difference is directly connected with the fact that the magnetic moment of the electron, which is equal to $eh/4\pi m$ (Bohr magneton), is roughly 1000 times greater than that of the proton or other nuclei. Nuclear moments according to equation (1) are principally given by $eh/4\pi m_P$ (nuclear magneton). As a result the two possible states for an electron (with a spin parallel or anti-

[11]) D. J. Kroon and C. van de Stolpe, Nature **183**, 944, 4th April 1959.

parallel to the external magnetic field) show an energy difference ΔE roughly 1000 times greater than that for a proton (see eq. (2)). Accordingly the frequency f_k at which nuclear magnetic resonance occurs (see eq. (4)) is given by:

$$f_k = g \frac{e}{4 \pi m_p} B,$$

and the frequency of paramagnetic resonance f_e (see eq. (5) in article [1])) is given by:

$$f_e = g_e \frac{e}{4 \pi m} B.$$

Since, for practical reasons, the magnetic fields B for observing the absorption due to electron spins and due to nuclear spins are chosen at the same order of magnitude (0.1 to 1 Wb/m^2), and since the numerical factors g and g_e in both formulae do not differ appreciably ($g_e \approx 2$, and g lies between 0.1 and 10, as mentioned earlier), we see that the resonance frequencies in the two methods differ by a factor of the order of 1000 (proton mass $m_p = 1836$ times electron mass m).

Another practical and very important difference between the two methods discussed concerns the sensitivity. The measurement of paramagnetic resonance is very much more sensitive than the measurement of nuclear magnetic resonance. This is again a direct consequence of the fact that for an electron the two possible states with opposite spin orientations show an energy difference ΔE about 1000 times greater than for a proton (confining ourselves now to protons). The relationship can be understood as follows. The difference between the numbers of protons and electrons in the two states mentioned is proportional to ΔE (see eq. (8), which is also valid for electrons). The number of electron spins that can contribute to energy absorption from the alternating field is therefore about 1000 times greater than the number of proton spins, the total number per unit volume being the same. Moreover the energy quantum ΔE absorbed from the alternating field upon each transition between the two states is also about 1000 times greater (this has already appeared in the frequency difference), so that the energy absorption — or the measured signal — other conditions being the same, is about 10^6 times greater for electrons than for protons.

True, the detectors used for centimetric waves give a much poorer signal-to-noise ratio than those for wavelengths of some tens of metres, and also the line width (to which the recorded intensity is inversely proportional) is several times larger in parametric resonance than in nuclear magnetic resonance. Nevertheless the first method referred to remains far more sensitive than the other, which means that paramagnetic-resonance investigations can be applied for very much smaller concentrations of resonators.

It is precisely for this reason that such useful results have been achieved with paramagnetic resonance: the paramagnetic centres are usually present in only low concentrations, whereas in nuclear-magnetic-resonance studies the whole of the investigated substance enters into the measurement.

This brings us to the last point of difference to be noted, which concerns the information obtainable with the two methods. Paramagnetic resonance provides information about the immediate environment of a paramagnetic centre. Nuclear magnetic resonance, on the other hand, gives information on the structure of molecules or of crystals as a whole, and on the effects of impurities (also non-paramagnetic impurities) on that structure.

A recent further development of the techniques discussed is the study of *double resonance*, where paramagnetic and nuclear magnetic resonance are excited in a substance simultaneously. This allows the nuclear magnetic resonance signal to be intensified, so that nuclei present in small concentrations are also accessible to investigation [12]). Moreover, the paramagnetic-resonance spectrum is influenced by the transitions between the energy levels of different nuclear spin orientations [13]). As a result still more particulars can be learned about the properties of the solid state. There is no space here, however, to deal at greater length with these methods and their application.

---

[12]) A. W. Overhauser, Polarization of nuclei in metals, Phys. Rev. **91**, 476, 1953. See also H. G. Beljers, L. van der Kint and J. S. van Wieringen, Overhauser effect in a free radical, Phys. Rev. **95**, 1683, 1954.

[13]) This method is also referred to as "Endor" (electron nuclear double resonance). See e.g. G. Feher, Electron spin resonance experiments on donors in silicon, Phys. Rev. **114**, 1219-1244, 1959 (No. 5).

Summary. Nuclear magnetic resonance gives rise to the specific absorption of electromagnetic waves (best measurable at metre wavelengths) by certain atomic nuclei, particularly hydrogen nuclei, when these are placed in a constant magnetic field. Since the resonance of a given nucleus is affected by the fields of neighbouring nuclei and electrons, the nuclear magnetic absorption spectrum offers a variety of information on the substance containing the nuclei. In solids the effect appears in a marked broadening of the resonance line; in liquids it appears in a splitting of the line into several components grouped closely together. A concise explanation of the theory of the phenomenon is followed by a description of two equipments used for recording the absorption spectra, one for liquids and the other for solids. The essentially different requirements of these two cases are dealt with at some length. In conclusion some examples are discussed of the numerous uses to which nuclear magnetic resonance has already been put, which include the identification of compounds (analysis of organic liquids) and the structural analysis of molecules and crystals

TEMPERATURES IN FITTINGS FOR INCANDESCENT LAMPS

by L. J. H. EXALTO.

536.4:628.95:683.865

For some years now there has been welcome evidence of increased activity in the design of lighting fittings, not only for the newer types of light source such as fluorescent lamps, but also for a source of long standing, the incandescent lamp. Presumably this activity has been stimulated by the post-war boom in housebuilding, by the evolution of public taste, and by the heightened interest of interior decorators in lighting matters. Decorators and others are experimenting with a variety of types of fitting, sometimes placing more emphasis on the aesthetic than on the functional aspect. Here it is for the public to decide whether they are prepared to sacrifice efficiency to appearance; but in any case safety must not be allowed to suffer. One aspect of the safety requirements relevant to incandescent-lamp fittings is dealt with in the article below.

Fittings for incandescent lamps must satisfy various safety requirements. In the first place the user must be safeguarded against accidently touching live parts. In the second place certain limits have to be imposed on the temperatures that arise under operating conditions. Obviously, a fitting that becomes excessively hot on the outside may represent a direct danger. If certain points inside the fitting become very hot there may be indirect danger too, in that conductors may short-circuit (an example is the point where the leads of the cord are splayed — point *s* in *fig. 1*) or come into contact with metal parts of the fitting and so make them live.

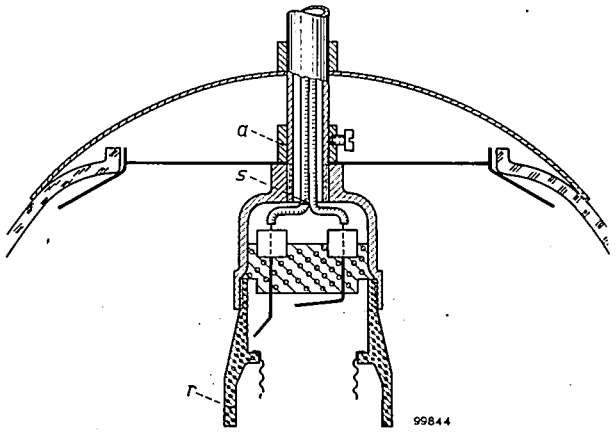


Fig. 1. Edison-type lampholder in fitting. The investigations involved the measurement of temperature of the lampholder rim (*r*), of the insulation of the cord at the point where the two conductors are splayed (*s*), of the ring containing the set-screw (*a*), and of various other points.

Details of safety requirements, and of test methods involving measurement of the temperatures arising in lamp fittings, are now being studied by the C.E.E. [1]). Fittings must, of course, be tested under conditions corresponding to heaviest loading in normal operation. This might mean inserting a lamp or lamps having the maximum wattage the fitting has been designed for. The matter is not as simple as all that, however, for the temperatures arising in a given fitting are not determined by lamp wattage alone, or in other words by the overall rate of heat production. They are dependent also on the temperature distribution within the lamp itself, which may be very different for different lamp types, as also upon the way in which heat is transferred to the fitting, i.e. upon the particular combination of conduction, convection and radiation operative therein: this last is dependent on the design and the shape of both lamp and fitting. All this makes it desirable that a standard heat source (called a "heat test source" by the C.E.E.) should be introduced for the purpose of testing fittings in a way that will allow their merits to be judged impartially.

The standard heat test source (to which we shall again refer at the end of this article) will have to be such that its heating effect is a reasonable approximation to that of a normal lamp. This aspect of the standard heat source has made it necessary to ascertain how various types of lamp behave in regard to the transfer of heat to the fitting. We shall now give some of the results yielded by extensive investigations undertaken by Philips in order to provide an answer to this question.

The first series of measurements we want to discuss relates to the simplest possible kind of "fitting" — a bare lamp suspended freely in still air from two copper wires. Of special interest here is the rise of temperature of the cap of the lamp above ambient. The rise in cap temperature may be attributed exclusively to conduction and to convection in the gas filling the bulb, processes which are influenced by the design of the filament and the dimensions of the lamp; radiation scarcely plays any part.

[1]) International Commission on Rules for the Approval of Electrical Equipment.

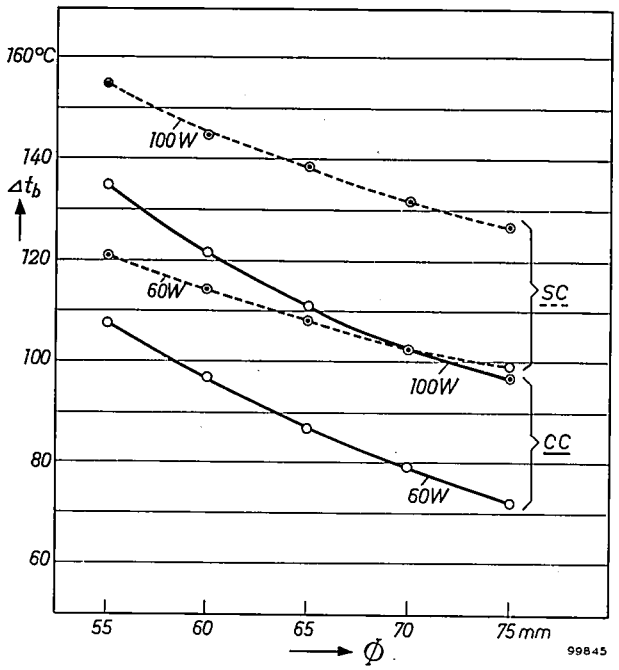


Fig. 2. The rise above ambient temperature of the rim of the cap of lighted incandescent lamps freely suspended from two copper wires in still air. This temperature rise Δt_b is plotted here as a function of Φ, the diameter of the glass bulb. The fully-drawn curves refer to 60 W and 100 W, 220-230 V lamps with coiled-coil filaments (CC). The broken curves refer to similar lamps with single-coil filaments (SC).

Some of the results of these measurements are displayed in *fig. 2*. The temperature rise of the rim of the cap, Δt_b, was measured [2]) in a series of gas-filled 220-230 V lamps of 60 W and 100 W ratings, some of which had single-coil and some coiled-coil filaments (indicated in the graph by the abbreviations SC and CC, respectively). Other things being equal, the amount of heat transferred to the gas filling is less in coiled-coil lamps than in single-coil ones [3]), and the former type could therefore be expected to produce lower values of Δt_b. To allow the influence of bulb size to be investigated, test lamps were made with diameters ranging from 55 mm to 75 mm in 5 mm steps, the other dimensions being adapted to the diameter (see *fig. 3*).

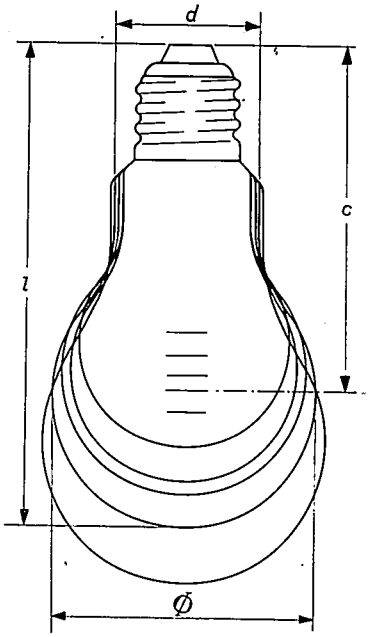
Fig. 2 shows that Δt_b increases as the lamp dimensions decrease. It also shows that, owing to the lower rate at which heat is transferred from filament to gas filling in a coiled-coil lamp, this type of lamp can be made with a bulb two or three steps of 5 mm smaller than that of a single-coil lamp of the same wattage, while retaining the same Δt_b value as the latter.

[2]) The method employed was that described in J. N. Bowtell, A method of measuring cap temperatures of general purpose tungsten filament lamps in free air, Light and Lighting **51**, 402-403, 1958.

[3]) This is precisely the reason why coiling the coiled filament results in higher luminous efficiency. See W. Geiss, Philips tech. Rev. **1**, 97, 1936 and **6**, 334, 1941.

Values of Δt_b considerably lower than those displayed in fig. 2 are found for vacuum lamps of the same wattage, likewise freely suspended in air. This is understandable, since there is no conductive or convective dissipation of heat via a gas filling; such dissipation in the gas-filled lamp has to be accepted in exchange for the higher filament temperature that is the favourable feature of the gas-filled as compared with the vacuum lamp (see for example the articles cited in footnote [3])). This point has been raised here in order to put the matter under discussion into true perspective: there is no case for abandoning the gas-filled lamp, with its far greater efficiency, just for the sake of reduced temperatures in fittings.

Having dealt with the bare lamp freely suspended in air, we shall go on to give results of temperature measurements on a fitting that may be regarded as an approximation to the other extreme. This would be a body closely enveloping the lamp and absorbing all the heat it gave off, including radiant heat. The temperatures of the various parts of a "fitting" of this kind would be entirely governed by the wattage of the lamp and, of course, by the provisions for cooling. The fitting actually investigated was made of metal, had a more or less conical shape, and was open at the bottom; no ventilation was provided for at the top of the fitting. The quantities measured [4]) included $\Delta t_s'$, the temperature rise at the cord splay point, and $\Delta t_d'$, the temperature rise in the metal



Φ	l	d	c
55	106	33	76
60	115	33	82
65	118	35	87
70	127	37	91
75	142	39	97

Fig. 3. Diagram showing dimensions of a series of incandescent lamps made for the purpose of the investigations. The lamps of the series had diameters Φ increasing in steps of 5 mm, the other dimensions being adapted to the diameter. The dimensions are given in mm in the table on the right. The horizontal lines lying roughly in the middle of the bulbs mark the position of the filaments. All the lamps had Edison-type E27/30 caps.

[4]) For a description of the method of measurement see H. F. Stephenson, The measurement of temperature in lighting equipment, Light and Lighting **47**, 369-372, 1954. (The same method is used by the C.E.E.) See also Flameproof electric lighting fittings, British Standard 889: 1947.

sleeve passing through the apex of the cone, the same series of lamps being investigated as in the case first discussed. In *fig. 4* these $\Delta t'$ values are plotted as a function of Δt_b, the temperature rise in the cap of the bare lamp, as given in fig. 2. It would seem that the conical fitting does not represent a very good approximation to the extreme case just alluded to (entirely enclosed fitting), since fig. 4 fails to demonstrate, except very roughly, that the temperature rises were determined only by the lamp wattage. Nor is another expectation borne out: it might be supposed that the temperature rise of the cap Δt_b would be representative of the heating effect of the lamp, i.e. that all $\Delta t'$ values for a given fitting would be mainly determined by Δt_b; it was to test this that we plotted our results in this particular way. If that supposition had been confirmed (and found valid for other forms of fitting), it would have been a comparatively simple matter to design standard heat sources for testing purposes. However, fig. 4 makes it quite clear that the facts are otherwise. Although there is unmistakably a correlation between Δt_b and $\Delta t_s'$ (say), it is a very incomplete one; for a given wattage the relative difference in $\Delta t_s'$ (or in the $\Delta t'$ value for some other point in the fitting) is very much less than the corresponding relative difference in Δt_b. This is clear from a comparison of lamps differing only in size, and it is even clearer from a comparison of lamps differing only in respect of filament design (CC and SC). Temperature rises in the fitting are thus but little affected by the considerable difference between the amount of heat dissipated via the gas filling in a CC lamp and that in a SC lamp.

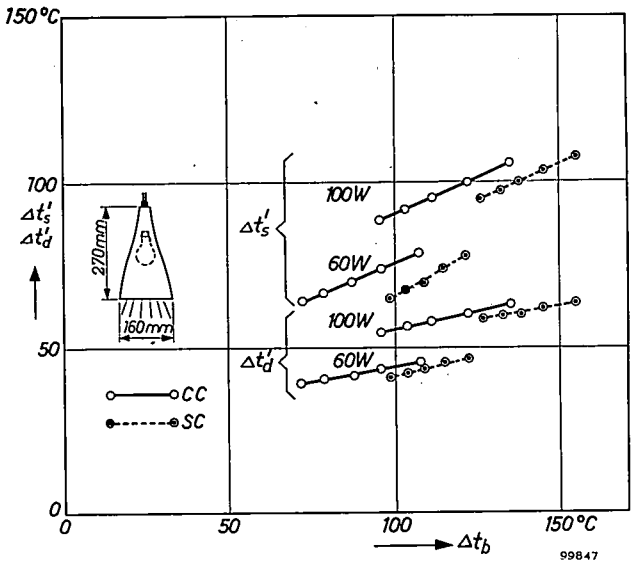


Fig. 4. Temperature rises measured with the same series of lamps as in fig. 2, inserted in a metal fitting of conical shape, open at the bottom and without ventilation at the top (see inset). The graph shows temperature rises $\Delta t_s'$ at the cord splay point, and $\Delta t_d'$ at the lead-in collar, as functions of Δt_b, the temperature rise undergone by the rim of the cap when the same lamp is freely suspended in air (fig. 2). As before, the fully-drawn curves relate to coiled-coil lamps and the broken curves to single-coil lamps. The five points through which each curve is drawn are the values measured (reading from left to right) for lamps with diameters of 75, 70, 65, 60 and 55 mm.
If the heating of the fitting were governed by lamp wattage alone, the curves would all be horizontal. If it were governed by Δt_b, the curves might be expected to pass approximately through the origin. The true state of affairs is evidently somewhere in between.

Finally, *fig. 5* displays the results of similar measurements performed on an entirely enclosed fitting having the form of an opal-glass globe. A fitting of this kind may be regarded as an intermediate case between the two extremes already dealt with — the bare lamp and the metal cone. The points selected for temperature measurements were the cord splay point ($\Delta t_s'$) and the rim of the lampholder ($\Delta t_r'$). Qualitatively, and quantitatively too, up to a point, results for this fitting bear a resemblance to those obtained with the metal cone. The relative differences in $\Delta t_s'$ (and $\Delta t_r'$) appropriate to different lamps are again considerably smaller than the corresponding relative differences in Δt_b. It was also found, as before, that the lamp wattage was by no means the only factor determining $\Delta t'$ values. However, one wattage rule does apply to both forms of fitting: replacing a 60 W by a 100 W lamp of the same design (i.e. with the same kind of filament — SC or CC — and the same dimensions) leads to an approximately 40% increase in Δt values. Furthermore, the results for the two forms of fitting corroborate the important conclusion that can be drawn from fig. 2,

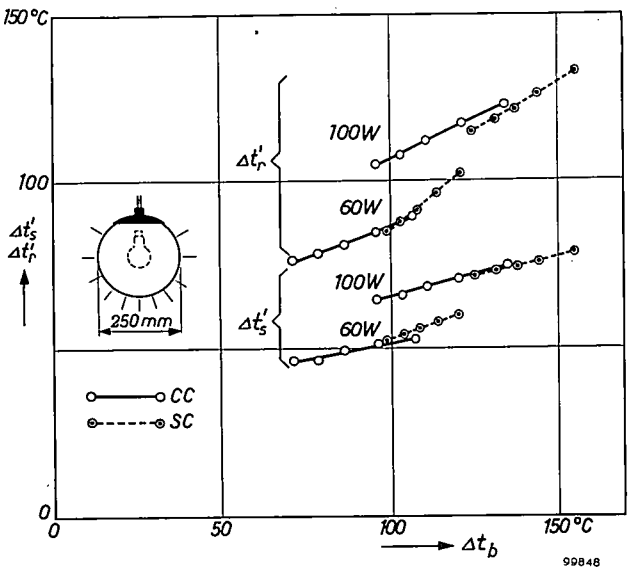


Fig. 5. Temperature rises measured inside an opal-glass globe (see inset) and plotted in the same way as in fig. 4. The temperature rises measured were $\Delta t_s'$, at the cord splay point, and $\Delta t_r'$, on the lampholder rim.

namely that, all other things being equal, smaller temperature rises are obtained with a coiled-coil filament than with a single-coil one. Hence, if the fitting has the form of a metal cone or an opal-glass globe, it is permissible (insofar at least as the points measured are concerned, which are, however, representative of the fitting as a whole) to use a CC lamp that is one or two steps of 5 mm smaller than a SC lamp.

The results given above — a very brief summary of extensive investigations, not yet concluded — provide no more than a rough guide in the approach to the problem of standard heat test sources. Even so, the results obtained so far are sufficiently revealing to justify a first attempt at making such sources. They will have to be provided with a normal lamp base, so that they can be inserted in fittings, and they will have to give off heat in much the same way as a normal incandescent lamp does; in other words, the sources will essentially be incandescent lamps (whose light yield, however, is of no interest). The experiments make it clear that it is not sufficient to specify certain Δt_b values: the wattage and mechanical design of the standard heat test sources, as well as their Δt_b values, will likewise have to be laid down in the specifications. A tentative specification for such heat sources is now in preparation. Practical tests will be necessary to assess their effectiveness.

Summary. For the testing of lighting fittings as regards their heating-up, it is desirable that standard heat test sources should be used in place of ordinary lamps. The problem is to design sources that will simulate, with reasonable exactness, the heating effect of incandescent lamps of the types that would normally be used in the fitting. Extensive investigations into this heating effect, carried out with lamps of different wattages and sizes in different fittings, have led to the conclusion that temperature rises are not governed entirely by the wattage or the Δt_b (temperature rise of cap of bare lamp freely suspended in air). Another important conclusion is that lamps with a coiled-coil filament cause smaller temperature rises in fittings than single-coil filament lamps that are otherwise completely similar.

A TRANSISTOR CARDIOTACHOMETER FOR CONTINUOUS MEASUREMENTS ON WORKING PERSONS

by G. A. HARTEN and A. K. KORONCAI.

53.083:612.16

Measurement of the heart rate is an important means of estimating the energy consumption of the human body and for ascertaining whether a person is subjected to needless physical strain when working under adverse conditions. The value assumed by the heart rate consists of a component related to the energy consumption of the body and of an extra component that may be due to a variety of factors, such as excessive ambient temperature, an awkward stance — certain muscles are then constantly contracted, which hinders the circulation in them — nervous tension, smoking and so on.

The relation between the first component and the energy consumption of the body can be determined in the laboratory in various ways with reasonable accuracy, and then correlated with the oxygen consumption, which is a direct measure of the energy consumed. Thus, if the heart rate and oxygen consumption are measured simultaneously, the magnitude of the extra component can be calculated [1]).

Omitting the oxygen-consumption measurement and determining only the heart rate does not, it is true, provide an exact indication of the energy consumption, but it still gives the physician a reasonable insight into the way in which the subject reacts to his work — unless the extra component is exceptionally large. Roughly speaking, rates of 70 to 80 beats per minute correspond to light (sedentary) work, 80-100 to moderately heavy work, and 100-130 to very heavy work. Of course, the terms light, heavy, etc., have no absolute significance here, but indicate how the subject himself (to judge from his heart rate) reacts to the work. When fatigue sets in, the extra component increases and so, therefore, does the heart rate.

The normal way of determining the heart rate by feeling the pulse is obviously not practical in an investigation of this kind; one cannot keep hold of a subject's pulse while he is working. If the subject suddenly stops what he is doing, the heart rate usually changes so quickly that a measurement taken immediately after the interruption already shows a quite considerable error, even when the pulse is taken for a period of only 10 seconds.

An apparatus has been developed at Eindhoven that makes it possible to measure the heart rate of working persons at any desired moment, and if necessary continuously, without obstructing their movements [2]). *Fig. 1* shows the equipment worn by the subject. The principle is as follows. Fixed to the subject's ear is a small clip, one side of which carries an electric bulb and the other a phototransistor type OCP 71. With each heart beat the blood stream

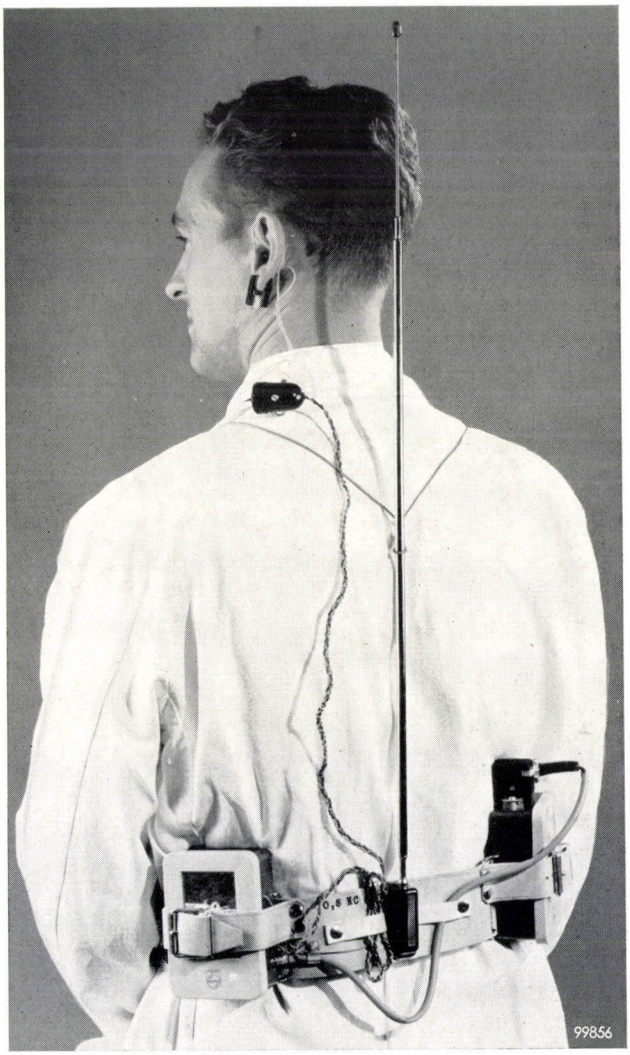


Fig. 1. The part of the apparatus worn by the subject. The subject is not obstructed in his movements.

[1]) See G. J. Fortuin, G. A. Harten, G. de Maar and P. A. van Wely, Tijdschr. soc. Geneesk. **37**, 389, 1959 (No. 10).

[2]) The same principle has been used by E. A. Müller and J. J. Reeh, Arbeitsphysiologie **14**, 137, 1950; E. A. Müller and W. Himmelmann, Int. Z. angew. Physiol. **16**, 400, 1957; and D. H. Bekkering and H. J. van Dal, Report No. 13, Research Inst. Public Health Eng., T.N.O., The Hague, June 1952. Literature referring to other equipment for cardiotachometry is quoted in the latter two papers and also in A. W. Melville and J. B. Cornwall, Electronic Engng. **31**, 268, 1959.

changes momentarily the amount of light transmitted through the ear lobe, and the current through the phototransistor therefore changes correspondingly. The resultant voltage pulses are conducted, after amplification, to a pulse shaper which delivers a square-wave signal upon every heart beat. This pulse signal triggers, via a transistor switch, a fixed-frequency oscillator (frequency about 3000 c/s) which amplitude-modulates the carrier frequency of a miniature transmitter (frequency 10 to 15 Mc/s). The signal from this transmitter can be picked up by a normal radio receiver.

Interference in the receiver due to other signal sources could be reduced if the carrier frequency of the transmitter were raised to about 100 Mc/s and frequency modulation adopted. An ordinary receiver could then still be used, but there would now be less interference from broadcast stations. Furthermore, the receiving aerial in this case could be directed towards the subject.

Each heart beat, then, is heard as a short high-pitched note, and the observer is able to count the number of beats occurring in a period of 10 or 15 seconds.

The pick-up element (*figs. 2* and *3*) and all other components mentioned, including the transmitter,

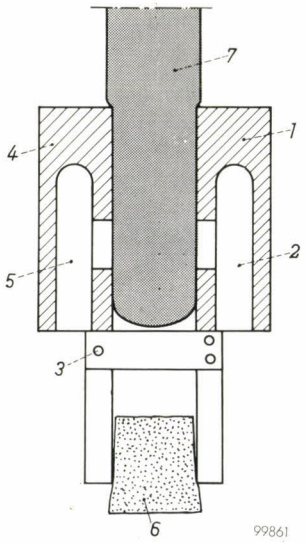


Fig. 2. Ear clip for electric bulb and phototransistor. *1* fixed arm with space *2* for bulb. *3* hinge pin. *4* hinged arm with space *5* for phototransistor. *6* foam plastic functioning as spring. *7* lobe of ear.

are very small and light and do not hamper the subject in his work. The way the apparatus is worn has been seen in fig. 1. Altogether the equipment weighs about three pounds. Except for the phototransistor, the entire electronic circuit (see *fig. 4*) is contained in the case on the right. The other case

Fig. 3. Showing how the clip is attached to the ear and how the wires are arranged so that they will not cause undue interference by communicating their movements to the clip. The clip itself weighs about $2\frac{1}{2}$ grams; the spring tension should not exceed 20 to 30 grams.

contains 18 small nickel-cadmium storage batteries and a potentiometer for adjusting the current through the electric bulb. The aerial is a normal telescopic type as used on cars. The apparatus can work for 8 hours before the batteries need to be recharged. This takes 10 to 12 hours.

To reduce interference, the amplifier is designed to pass only frequencies in a band of about 1 to 3 c/s, i.e. 60 to 180 heart beats per minute (see *fig. 5*). The chief causes of interference are the mains (hum) and unduly sharp movements of the subject. Interference signals of the latter kind have a frequency of the order of 10 c/s. If there is a powerful source of infra-red radiation near the subject, the ear must be screened; infra-red and also visible red radiation are passed fairly well by the ear, and the phototransistor reacts to radiation up to wavelengths of about 1.5 μ [3]).

If frequent or continuous measurements are called for, simple counting of the audible pulses by an observer is evidently inconvenient. We have therefore designed a recording count-rate meter for the

[3]) See e.g. F. H. R. Almer and P. G. van Zanten, An experimental pyrometer using a phototransistor and designed for radio-tube inspection, Philips tech. Rev. **20**, 89-93, 1958/59 (No. 4).

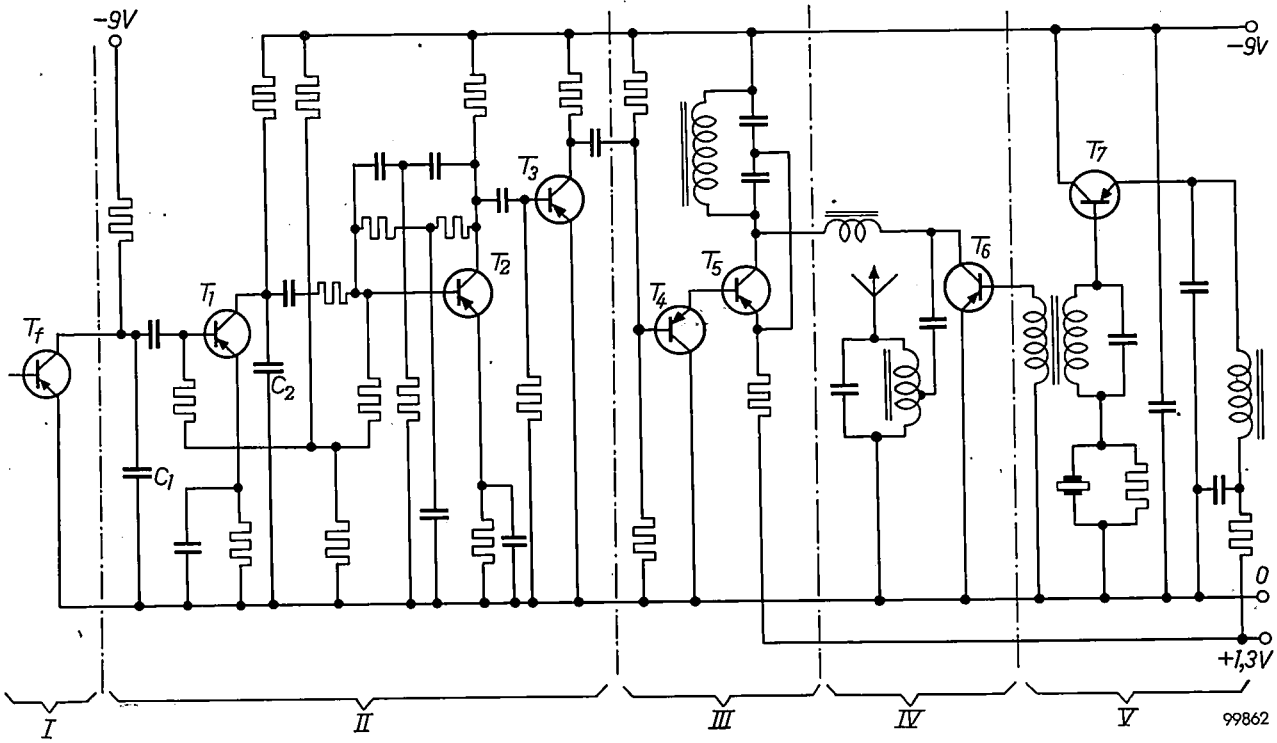


Fig. 4. Electronic circuit of apparatus carried by subject. On the left, in section *I*, is the phototransistor T_f, which is mounted in the ear clip. Section *II* contains the selective amplifier and pulse shaper. The required frequency response is obtained by means of a double T-filter [4]) and capacitors C_1 and C_2 (1.6 μF). Section *III* contains a 3000 c/s oscillator (transistor T_5), triggered into oscillation by a switching transistor T_4 while the shaper circuit in *II* delivers a pulse. This oscillator modulates the carrier frequency (10-15 Mc/s) of the transmitter. The latter consists of a crystal-controlled transistor oscillator (section *V*) inductively coupled to an output stage (section *IV*). The phototransistor T_f is fed from a separate voltage source. The circuit in section *II* is encapsulated in a plastic block measuring 35 × 35 × 70 mm. Sections *III* and *IV* are together in a similar block measuring 30 × 35 × 90 mm, and section *V* in another block of the same size.

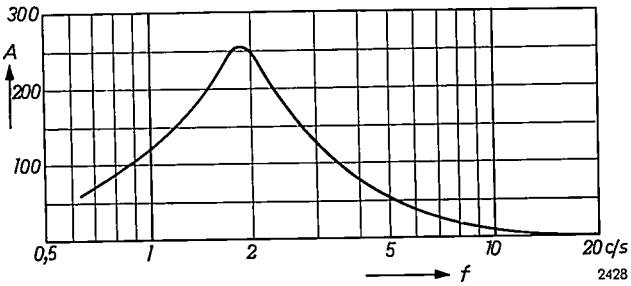


Fig. 5. Frequency response (voltage gain A versus frequency f) of amplifier (section *II* in fig. 4). At frequencies outside the range from 1 to 3 c/s the gain is less than half the maximum.

purpose. A block diagram of the instrument is given in *fig. 6*. Block *1* is the radio receiver, which gives a short-lived signal of 3000 c/s at every heart beat. This signal passes to block *2*, in which it is fed to a selective amplifier that only passes frequencies in a narrow band at 3000 c/s (to limit interference), and is then converted into a rectangular pulse. Block *3* contains a count-rate circuit consisting of a monostable flip-flop followed by a diode pump circuit [5]) and an *RC* network in which a meter *4* is incorporated. The time constant of the *RC* network is approximately 5 seconds. The meter indication is thus in a sense an average of the heart rates that occurred in the 5 to 10 seconds prior to the moment

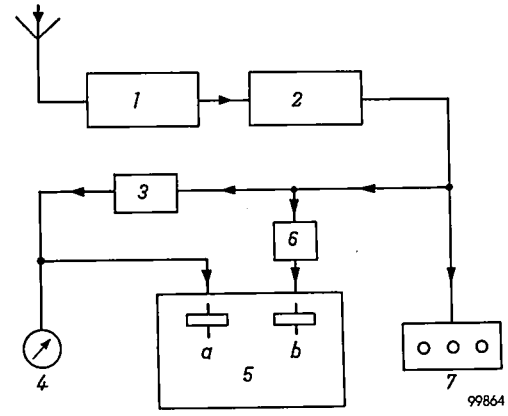


Fig. 6. Simplified block diagram of apparatus for measuring and recording the heart rate. *1* radio receiver. *2* selective amplifier (3000 c/s) with pulse shaper. *3* count-rate circuit. *4* meter with scale from 1 to 3 c/s (60 to 180 beats per minute). *5* recorder; the stylus *a*, like the meter *4*, receives the output signal from the count-rate circuit. *6* circuit enabling the stylus *b* of *5* to record *rapid* variations of the heart rate. *7* electrical counter.

[4]) The properties of these filters are described in D. H. Smith, Electronic Engng. **29**, 71, 1957.

[5]) The diode pump circuit is dealt with extensively in J. B. Earnshaw, Electronic Engng. **28**, 26, 1956. A recent application is described in J. J. van Zolingen, Philips tech. Rev. **21**, 134-144, 1959/60 (No. 4/5), on page 138 *et seq.*

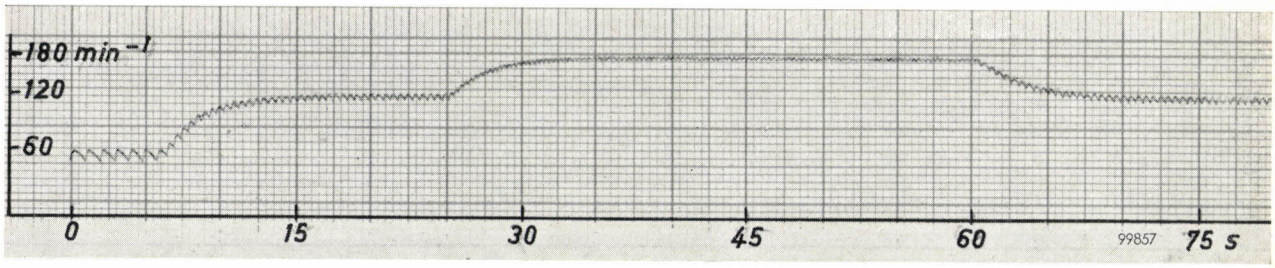
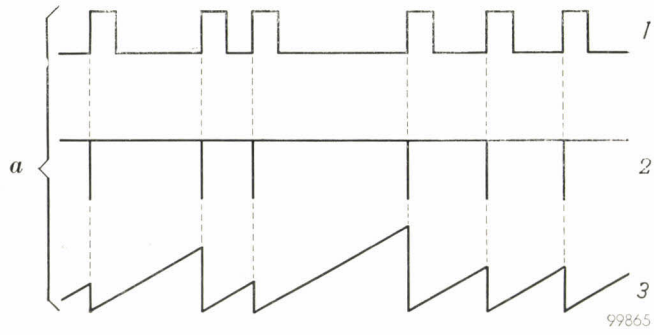


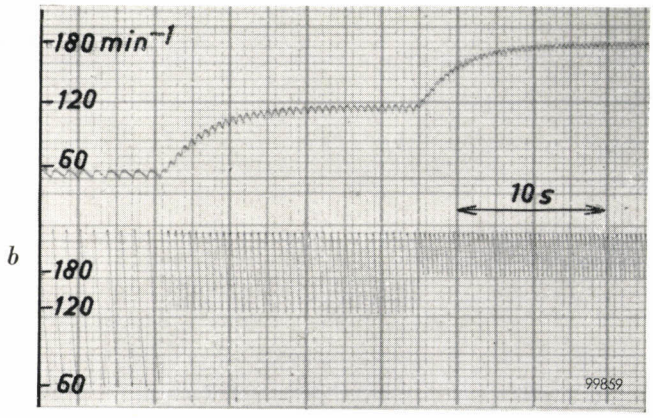
Fig. 7. Recording of the output signal from the count-rate circuit (block *3* in fig. 6) at frequencies of 1, 2 and 3 c/s (60, 120 and 180 beats per minute). The recording stylus drops back slightly between two beats. The response time here is about 12 seconds.

of taking the reading. Of course, the meter cannot follow very rapid changes of heart rate, but the response time is short enough for most cases encountered. If the *RC* time constant is made shorter, the needle deflection drops appreciably between successive heart beats, and consequently the indication is more difficult to read. The value mentioned proved to be a useful compromise between the readability of the meter and its response to the variations of the heart rate. The circuit is designed to allow the meter scale to cover a frequency range from 1 to 3 c/s. The output signal of block *3* can also be recorded by the stylus *a* of the recorder *5* (see *fig. 7*).

For following *rapid* variations of the heart rate, use can be made of the circuit *6* in fig. 6. This delivers sawtooth pulses whose amplitude is a measure of the time elapsing between two heart beats (*fig. 8a* and *b*). When these pulses are recorded and a line drawn through their peaks, the resulting curve is effectively that which would be obtained by a recording frequency-meter having a very short indication time [6]). The recorder itself will produce this curve directly if the paper is made to move slowly enough (fig. 8*c*). The paper speed should not, of course, be so reduced as to lose the advantage of rapid indication.

Block *7* in fig. 6 represents an electrical counter, equipped with cold-cathode tubes, that counts all pulses (heart beats) and can therefore be used when the observer wishes to make a direct count of heart beats over a given period.





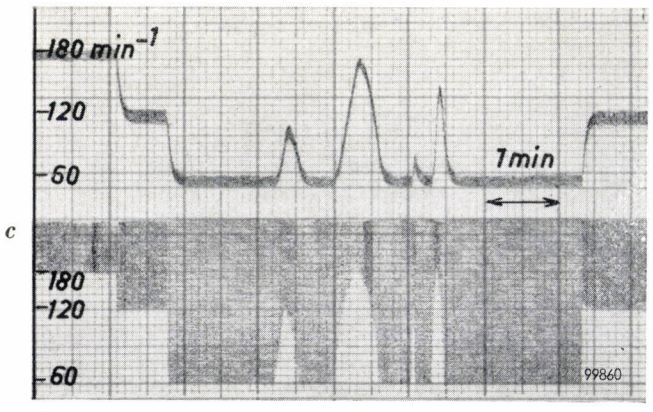


Fig. 8. *a*) The circuit in block *6*, fig. 6, operates in the following way. The output pulses *1* from block *2* (obtained by rectifying the 3000 c/s signals) are applied to a pulse shaper that converts them into pulses *2* of much shorter duration (about 3 ms). These pulses, via a transistor circuit, discharge a capacitor, which charges up again in the time between two pulses. The voltage across this capacitor shows the waveform *3*. The height of the sawtooth voltage peaks is a measure of the time interval between successive heart beats.
b) Above: recording of the output signal from the count-rate circuit. Below: recorded output signal of circuit mentioned under *a*). The applied frequencies were successively 1, 2 and 3 c/s in both cases. A line drawn through the peaks (the envelope) of the lower trace can be regarded as originating from a count-rate meter having a very short response time.
c) The same curves recorded in a practical case. By reducing the speed of the paper the envelope referred to is obtained automatically in the form of the demarcation line between the written and unwritten parts of the paper.

[6]) The idea of applying this method to the present case is due to G. Klein of this laboratory.

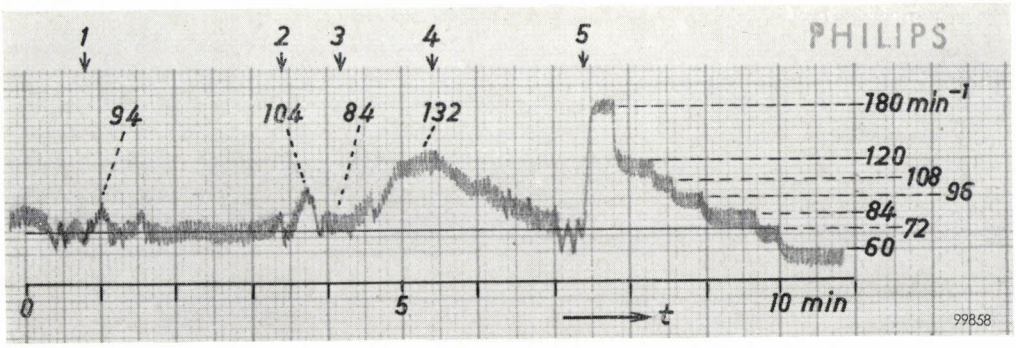


Fig. 9. Recording of heart rate versus time t (left, up to the point 5). Initially the subject is seated. At *1* he shifted his chair for a moment and then remained quietly seated. At *2* he stood up and remained standing; about 25 seconds after the action of standing up, the heart rate has returned to a virtually constant but somewhat higher value than when the subject was seated. Between *3* and *4* the subject performed twenty deep knees-bend exercises and then sat down again (at *4*). After 2 minutes the heart rate has returned to the normal value when seated. On the right, from *5* onwards, the recording was made from an "artificial heart" to obtain the calibration marks shown.

If the subject does not have to shift from one place to another during his work, and need make no considerable movements, wireless communication is evidently superfluous. The output signal from the pulse shaper that drives the 3000 c/s generator can then be applied directly to the measuring and recording instrument (fig. 6, blocks *3* to *7*). This does away with the interference produced by the radio receiver picking up signals of different origin. Moreover, the subject then carries no other equipment than the ear clip and leads. The other apparatus can be set down beside the count-rate meter and recorder.

An example of the way the heart rate reacts to various kinds of bodily movement is given by the recording in *fig. 9*. The calibration marks on the graph were obtained with the aid of an "artificial heart", consisting of a metronome to which a small plate is attached. With each swing of the metronome the plate interrupts the path of light between a lamp and a phototransistor. This device also supplied the pulses required for recording figs. 7 and 8.

Summary. Investigations into the physiological aspects of work call for a means of continuously measuring the heart rate of persons engaged on their normal work, where it is evidently not feasible to take the pulse in the normal way. In the apparatus described the individual heart beats are converted into electrical pulses by arranging for a beam of light transmitted through the lobe of the ear to fall on a phototransistor. Each pulse causes a miniature transistorized transmitter to transmit a note of 3000 c/s on a carrier of 10-15 Mc/s. Together with aerial and Ni-Cd batteries, this part of the equipment weighs about 3 lb and is carried on the back. The transmitted signals are picked up by a conventional radio receiver and either counted directly or applied to a recording count-rate meter (range 1-3 c/s; time constant ~ 5 sec). Rapid changes in heart rate can be faithfully recorded by means of a special circuit that supplies pulses whose amplitude is a measure of the time between two heart beats.

AUTOMATIC CONTROL OF A FILAMENT-COILING MACHINE WITH THE AID OF PRESET COUNTERS

by F. EINRAMHOF and P. HAVAS.

621.326.652.3:621.778.3-52

In manufacturing certain types of filaments for incandescent lamps the filament-coiling machine is usually controlled by a cam mechanism. This method is not particularly convenient when it is desired from time to time to switch the production to different types of filament. This article describes an electronic control circuit incorporating cold-cathode tubes, in which the lengths of the coil sections can be adjusted by means of switches.

Special-purpose incandescent lamps, e.g. for studio lighting, film projection and coastal navigation lights normally have filaments composed of a number of coiled sections spaced by sections of uncoiled wire (*fig. 1*).

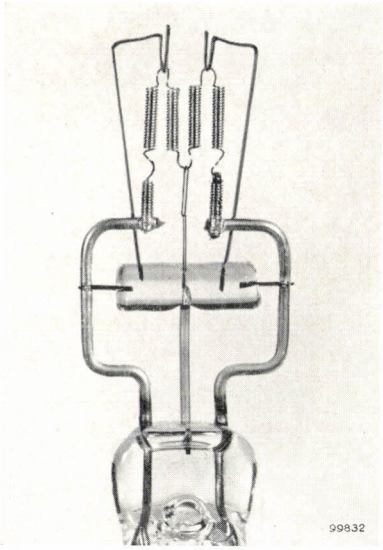


Fig. 1. Filament composed of four parallel coiled sections (See Th. J. J. A. Manders, Incandescent lamps for film projection, Philips tech. Rev. **8**, 72-81, 1946, page 74.)

There are two methods of manufacturing these filaments. In the first method the wire is coiled into a continuous helix. Afterwards the straight sections are made by straightening one turn at the appropriate location. If the straight sections have to be longer than the length of a single turn, the second method is adopted. Here the straight parts are produced automatically: the winding machine produces a coil interrupted by straight sections of wire.

For the second method the winding machine must be controlled according to a certain programme, which was hitherto realized mechanically by means of cams. For changing over to the manufacture of another type of filament, certain gears in the camshaft drive have to be exchanged. For large runs this can hardly be called a drawback, but for small runs a more flexible form of control is preferable. It is all the more desirable for development work on filaments for new applications or for improvements to existing types. One important problem is attaining the optimum temperature distribution within the filament. For systematically investigating this matter, the lengths of coil sections and gaps must be varied independently. With cam-motion control this would become a costly and elaborate operation.

An electronic device using preset decade counters, i.e. counters giving a voltage pulse when reaching a preset number, has now been designed for the control of filament-coiling machines. The preselected programme can be easily and rapidly changed by means of a few switches. An experimental set-up of a simple filament-coiling machine controlled by this circuit will be discussed here.

Description of the coiling machine

A sketch of the coiling machine is shown in *fig. 2*. A bobbin *S* with tungsten wire is fitted to winding

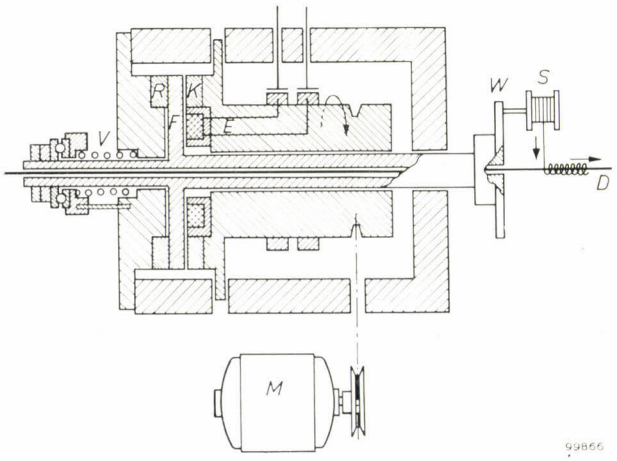


Fig. 2. Sketch of the winding machine. *M* motor. *W* winding head. *S* bobbin with tungsten wire to be wound on mandrel *D*. *F* flange fitted to the winding-head spindle. *E* electromagnet, which, when energized, pulls the flange against clutch plate *K*, so that the winding head is driven by the motor. *V* spring, pressing the flange to braking disc *R* when the electromagnet is not energized, thus stopping the winding head.

head W. When electromagnet E is energized, flange F is pulled to the right against clutch plate K; the winding head is now driven by the motor. When E is no longer energized, the disc is pressed by spring V to braking disc R and the winding head is stopped. The wire is coiled around a mandrel D moved axially through the winding-head spindle by the motor. (This mandrel is subsequently dissolved in acid, so that only the coil remains.)

Winding a filament composed of, e.g., three coiled sections and three straight parts is done in the following stages.

a) The motor is running and the electromagnet is energized, hence the mandrel is moving and the winding head is rotating, so that the first coiled section is being wound.

To stop winding, the current through the electromagnet must be interrupted. If this were to happen with the motor running at full speed, some slip would occur because the brake cannot immediately bring the head to a standstill. The pitch of the coil would therefore increase and the coil would be too long. Besides, the slip of the head is not always uniform, so that deviations from the nominal lengths cannot be prevented by interrupting the current at an earlier moment. The speed of the winding head must therefore be reduced first. This comprises the next stage, viz.:

b) The motor is switched off, whilst the electromagnet remains energized. The motor is running out but winding is still continuing at decreasing speed.

When the speed of the winding head has dropped sufficiently, the electromagnet is cut out, stopping the winding head. At the same time the engine is revved up again for the third stage.

c) The motor is running, but the electromagnet is not energized. The mandrel is now moving but no winding takes place, so that a straight section of the filament is being formed.

The coil is thus made in stages (*a*) and (*b*), the straight part in (*c*). For the next coil-and-straight section these three stages are repeated, so that a filament comprising three such sections is made in nine stages. The velocities of motor spindle and winding head during four successive stages are shown in *fig. 3*.

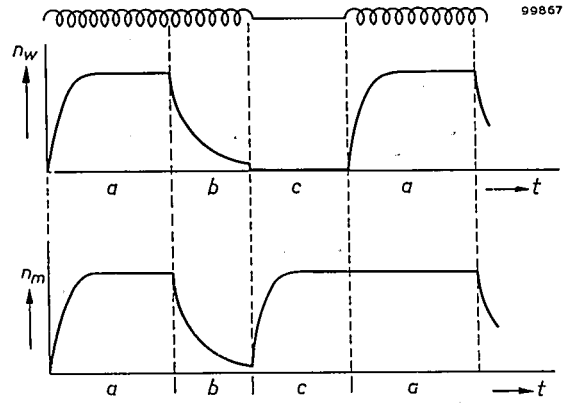


Fig. 3. Rotation speeds n_m of the motor spindle and n_w of the winding head as functions of time: *a* coil winding at full speed, *b* winding at decreasing speed, *c* formation of straight section of filament. The coil thus produced is shown schematically above the diagrams.

Principle of the control circuit

Each stage must be switched on as soon as the preceding stage is complete, i.e. as soon as the coil has the necessary number of turns or the straight section has the determined length. This implies, in each case, that the motor spindle has then completed a certain number of revolutions. If the entire process is to take place automatically, a device is required to count the number of revolutions of the motor spindle, and to produce a switching pulse when this number is reached. These numbers must be readily reset to make it possible to vary the lengths of the coils and straight parts independently of each other.

For counting the number of revolutions of the motor spindle the latter is fitted with a disc P with ten apertures (*fig. 4*). On one side of the disc is a lamp L, on the other side a phototransistor Tr which triggers a transistorized pulse generator I ten times per revolution. The pulses from I (height 80 V, length 25 μsec) are applied to a decade counter Te, consisting of four decade-stages of cold-cathode tubes. The maximum counting capacity is 9999 pulses, or a 1000-turn coil, with an accuracy of 1/10 turn.

The counter is connected to a programmer unit PB, likewise comprising cold-cathode tubes. When the preset number of pulses has been counted, the programmer applies a signal to switching unit SB upon which the latter energizes motor and/or electromagnet for the next stage. At the same time the counter is reset to zero, so that the number of pulses of the next stage can be counted.

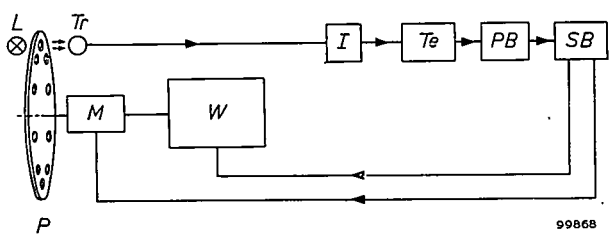


Fig. 4. Block diagram of the winding machine with electronic control. W winding machine. M motor. P perforated disc on motor spindle. L light source. Tr phototransistor. I pulse generator. Te counter. PB programmer unit. SB switching unit.

The counter circuit described here thus gives a switching command when the preset number is reached, unlike decade counter tubes such as the E 1 T [1]), which only *indicate* the number of pulses counted.

The counter

The cold-cathode gas-discharge tubes employed are of the type Z 70 U (*figs 5a* and *b*). Apart from cathode and anode, this tube contains two auxiliary electrodes, viz. a trigger electrode and priming electrode arranged as second cathode. This latter

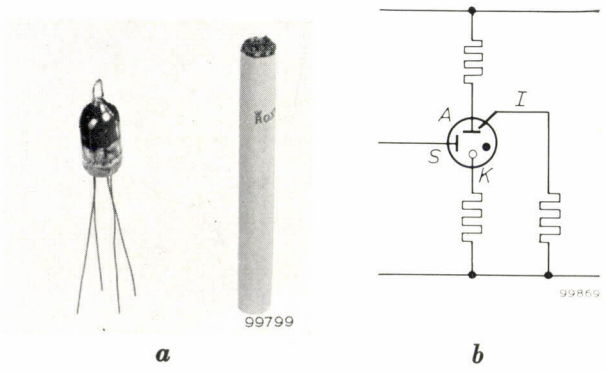


Fig. 5. *a*) Cold-cathode gas-discharge tube, type Z 70 U. *b*) Schematic diagram. *A* anode. *K* cathode. *S* trigger electrode. *I* priming electrode. A small discharge current is permanently present between priming electrode and anode, to minimize the delay in establishing the main discharge.

electrode is fitted very close to the anode, so that a priming discharge is initiated by an extremely small current (3 μA); to limit the priming current to this small value the priming electrode is earthed via a high resistance. The priming discharge ensures that ions are always present within the tube, so that the main discharge is established more quickly. The trigger voltage (trigger-cathode) to initiate the main discharge is 145 V; the burning voltage between cathode and anode is 120 V.

Each decade stage comprises 10 tubes, circuited as shown in *fig. 6*. The +80 V pulses to be counted are applied to input *I*. Consider the situation when only valve *0* is conducting. This tube then has a cathode potential of 100 V; the trigger electrode of tube *1* has the same potential. When the next counting pulse reaches the input, the triggering voltage of tube *1* is exceeded, so that now this tube ignites. The pulse is likewise supplied to the triggers of the other tubes, but these do not ignite as they have as yet no trigger bias.

When tube *1* ignites, tube *0* extinguishes. The cathode of tube *1* initially remains at earth potential owing to the presence of C_{k1}, so that a large current flows through tube *1*. This current temporarily sets up a high voltage across the common anode resistor R_a, so that the anode voltage of all tubes drops. The cathode of tube *0*, however, is still maintained at 100 V by C_{k0}, so that the anode-cathode voltage of this tube drops below the burning voltage. Whilst tube *0* extinguishes, tube *1* remains conductive, because the cathode potential of the latter tube is as yet low, so that here the anode-cathode voltage does *not* fall below the burning voltage. While the voltage across C_{k1} rises towards 100 V, the current through tube *1* decreases.

Because tube *1* is now conducting, a 100 V bias is applied to the trigger of tube *2*, so that upon the next pulse the latter tube ignites and tube *1* extinguishes. Each successive tube is similarly ignited. The number of the tube conducting at any given moment thus indicates the total number of pulses received.

The maximum counting rate is determined by two factors. Once a tube has been ignited by a counting pulse, the next counting pulse must not arrive before the cathode voltage of this tube (and hence the bias of the next tube) has reached 100 V. It is accordingly the *RC*-time of the cathode circuit that determines the counting rate. This is moreover restricted by the time necessary for initiating the discharge. The present circuit has a maximum counting rate of 2000 pulses/sec.

When tube *9* is conducting, the next pulse has to initiate two events. First it has to ignite tube *0* of its own decade. This means that tube *0* must receive its bias from tube *9*. The ten counter tubes are accordingly arranged as a ring circuit. Secondly a pulse must be applied to the next decade. That is why the first decade is equipped with an additional tube, the "relay" tube *D* (fig. 6), arranged in a self-quenching circuit. *D* is biased when tube *9* is conducting, and will be fired by the next counting pulse. The relay tube, however, is not directly connected to the common anode resistor of the decade, but via a second resistor R_a' of sufficiently high value to prevent a sustained discharge within the tube. A capacitor C_a is connected in parallel with the anode resistors, allowing the tube to fire for a moment, but quenching it as soon as the voltage across this capacitor has been built up. This means that the cathode voltage just reaches 100 V for a moment and then drops to zero. The latter pulse is trans-

[1]) See the article: A. J. W. M. van Overbeek, J. L. H. Jonker and K. Rodenhuis, A decade counter tube for high counting rates, Philips tech. Rev. **14**, 313-326, 1952/53.

The preset decade counter employed and the programmer-tube circuits described below were developed by L. Wasser and E. Strauman of Philips AG Zurich.

A similar counting circuit for an entirely different application was described some years ago: J. Domburg and W. Six, A cold-cathode gas-discharge tube as a switching element in automatic telephony, Philips tech. Rev. **15**, 265-280, 1953/54.

mitted from the output O to the input of the next decade.

In this way the four decades are connected in cascade. The indication of the counter is given by the trigger tubes themselves which light up when ignited, behind windows with the appropriate numbers.

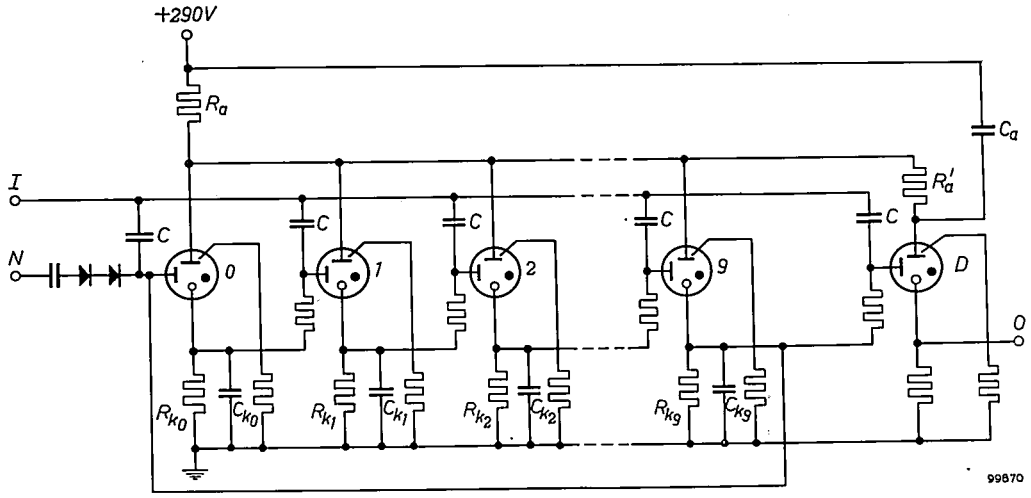


Fig. 6. Decade of a decade counting circuit employing cold-cathode tubes (Z 70 U). *0-9* counting tubes. *D* relay tube. *I* input. *C* input capacitors. R_a common anode resistor. R_a' additional anode resistor for tube *D*. C_a reservoir capacitor whereby *D* passes a short current pulse when fired. *O* output to the next decade. *N* input for the reset pulse, for returning the counter to zero.

For resetting the counter to zero, a +300 V pulse is applied (at N) to tube *0* of all four decades. This is sufficient to ignite these tubes even if they have no bias, whilst the temporary drop in anode voltage extinguishes all other tubes. Reset to zero is commanded by the programmer unit, which produces a 100 V voltage pulse for this purpose. This voltage is applied to the zero-reset tube, a cold-cathode tube supplied with a permanent bias from a voltage divider. Like the relay tube of a decade, it is arranged in a self-quenching circuit and therefore produces only a pulse of current. This current is stepped up to the required 300 V by a transformer.

The programmer unit

The programmer unit also incorporates cold-cathode tubes of type Z 70 U, viz. one tube per decade for each stage of the coiling process. The circuitry of such a programmer tube is shown in *fig. 7*. The tube gets its trigger bias from another programmer tube. The trigger of the tube is connected, via an input capacitor C and a 10-position switch, to the cathode of one of the counting tubes of a decade. When the decade reaches the preselected number on the switch (except zero, see below), a voltage of +100 V relative to earth is applied to C. The charging current of this capacitor causes a voltage pulse across resistor R that will fire the tube if the trigger electrode has the necessary bias.

The operation of the complete circuit of counter and programmer tubes may now be explained with the aid of the block diagram of *fig. 8*. Like the tubes of each counting decade, the programme tubes P_1-P_{36} are arranged in a ring. The tubes P_1-P_4 correspond to the number indicating the desired length of the first coiled section of the filament; let us assume that in the counter the number 6527 has

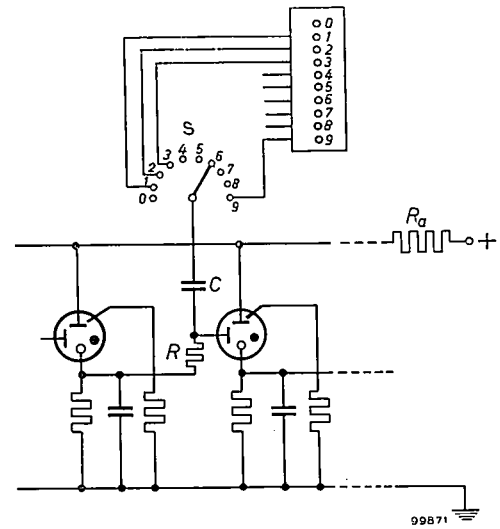


Fig. 7. Cold-cathode tube as programmer tube. The ten-position switch S connects the trigger electrode of the tube to the cathode of one of the counting tubes of the decade. The tube is biased via the cathode of the preceding programmer tube and then fires when the counter indicates the appropriate number. The programmer tubes have a common anode resistor R_a.

been preselected. This means that the trigger of P_1 is connected to the cathode of tube *6* of the thousands decade, P_2 to tube *5* of the hundreds decade, P_3 to tube *2* of the tens decade and P_4 to tube *7* of the units decade. Before the winding process is started, the counter is set to zero and the bias is applied to P_1. The motor is now started, so that count pulses are applied to the input of the counter. When the thousands decade changes to *6*, P_1 ignites. This produces the bias for P_2. Now the next occasion that the hundreds decade reaches *5*, P_2 can fire. 6500 pulses have now been counted. P_1 extinguishes again and P_3 is biased. Then, after 6520 pulses, P_3 will fire and similarly, after 6527 pulses, P_4.

The cathode of P_4 then has a voltage of 100 V. This is used to start the next stage as follows.

1) The voltage is transmitted to a switching unit, which changes the activation of motor and magnet as required for ending the first stage in the winding process and starting the second stage.
2) The voltage is applied to the zero reset. The counting decades are thus set to zero and are immediately ready for counting the second number of pulses.
3) P_5 is biased.

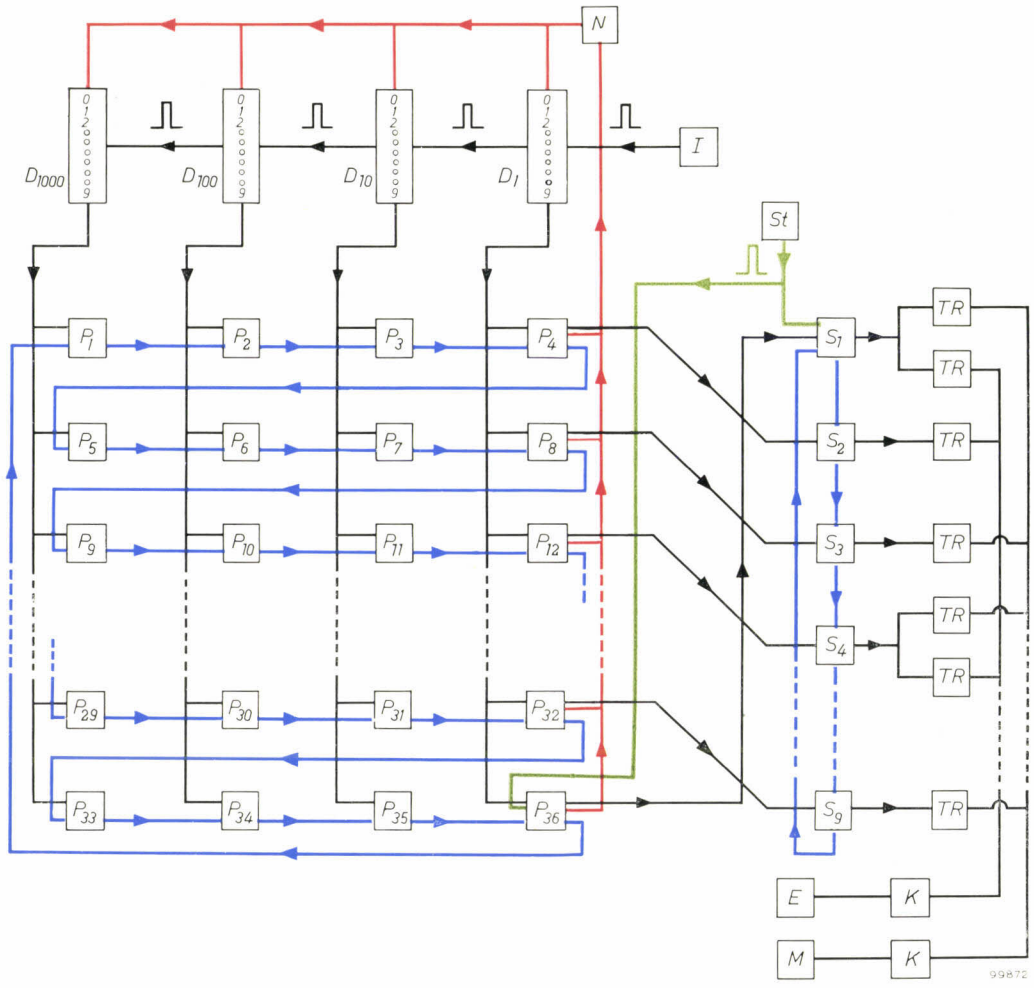


Fig. 8. Block diagram of the control circuit. D_1, D_{10}, D_{100}, D_{1000} decades of the counter. *I* pulse generator. *N* zero resetter. *St* starting circuit. P_1-P_{36} programmer tubes. S_1-S_9 switching tubes. *TR* Schmitt trigger circuits. *K* power transistors. *M* motor. *E* electromagnetic clutch. Black connections are those for the count pulses, firing voltages and switching voltages; blue connections are for the bias voltages; red connections are for the zero reset voltage; green connections are for the starting pulse.

Whilst the second stage is being completed, valves P_5-P_8 are successively ignited, after which the third stage is started. In this way the entire programme is completed. The last tube of the ninth stage applies again the bias to P_1, so that production of the next filament starts without interruption.

In position *0* the ten-position switches have a function different from that described above. Con-

sider the case in which 6501 pulses have to be counted for the first stage. P_1 will then normally fire at 6000, P_2 at 6500. For 6501, P_3 should fire simultaneously with P_2 and supply the bias to P_4, since that tube has to fire at the *next* count pulse. If the trigger of P_3 (with the corresponding switch set to position *0*) were connected to the cathode of tube *0* of the tens decade, however (as corresponds to other positions of the switch), P_3 would not fire at all. This is because after 6499 pulses it is *first* the unit decade of the counter that changes to zero and then the tens decade, and only then does the latter supply a pulse to the hundreds decade, changing it to 5. This fires P_2 and only then will P_3 be biased. P_3, however, can no longer fire, because tube *0* of the tens decade is already conducting and the ignition pulse is no longer available. In this case it is accordingly necessary to bypass P_3 in the circuit; P_4 must obtain its bias directly from P_2. It is therefore arranged that when a ten-position switch is in position *0*, the bias received from the preceding programmer tube is directly relayed to the next programmer tube.

The switching unit

Next to the programme tubes in fig. 8 are shown the switching tubes S_1-S_9, one for each stage of the winding process. These nine tubes are similarly arranged in ring configuration. The cathode of each tube is connected to either one or two Schmitt trigger circuits, which are triggered whenever the corresponding programmer tube ignites, thereby controlling the current to the winding motor and/or magnetic clutch via power transistors.

To start the whole winding process, a 300 V pulse is required. This pulse is obtained by discharging a capacitor across a resistor when a push-button is depressed. The pulse ignites S_1 and P_{36}. The two Schmitt circuits connected to S_1 are then triggered, energizing motor and magnet for the first stage (coiling). The cathode voltage of P_{36} has three effects: 1) zero-resetter N is activated; 2) P_1 is biased and the first number is now counted off as described above for the programmer unit; 3) the voltage is applied to the trigger electrode of S_1 (which is necessary at the end of the programme). Because S_1 is now conducting, S_2 is biased. The cathode of P_4 is connected to the trigger electrode of S_2. When the first number is reached, P_4 fires and S_2 fires immediately afterwards. This extinguishes S_1, and motor and magnet are cut out. The Schmitt circuit following S_2, however, is immediately triggered, thus directly energizing the magnet again. The magnetic clutch does not respond to this very short interruption of the current. The motor remains switched off, and winding is slowing down. P_4 has meanwhile also activated the zero-resetter and biased P_5, so that now the second number is counted off. Upon completion, P_8 ignites, so that S_3 (biased by S_2) fires: the third stage therefore commences (motor only switched on, forming straight section of filament). During the last stage, S_1 is biased and can therefore be ignited by P_{36}, enabling the programme to be repeated without interruption, i.e. to wind a second filament, and so on. During each stage the corresponding switching tube remains conducting, so that there is always visual indication of the stage in progress at any moment.

By means of the 36 ten-position switches the 9 numbers can be independently varied, which makes a highly flexible system.

Construction of the equipment

The major part of the circuit is accommodated in boxes of $40 \times 104 \times 117$ mm. Printed wiring is used for all component parts, and connections between containers are established by contact strips forming part of the printed wiring. The contact strips fit into special sockets arranged on a panel. Each decade is housed in its own container, together with two programmer tubes and the two corresponding digit switches (*fig. 9*). The position of the counter is indicated through the windows at the front. The

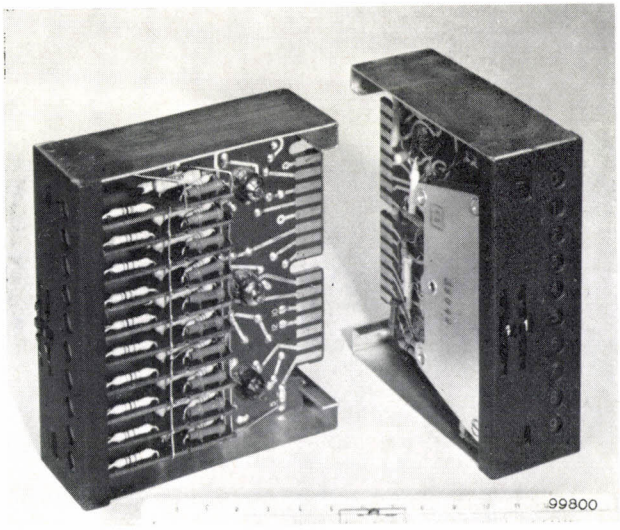


Fig. 9. Two views of a decade of the counting circuit; each decade chassis also houses two programmer tubes and their ten-position switches. If more than two stages are to be programmed (nine in the case described here), additional programmer tubes are available (for each decade). The numbered windows behind which the counting tubes are fitted can be seen on the front of the decade boxes. The tubes are painted black to prevent premature firing by incident light. The decade switches together with figures indicating their positions can also be distinguished.

digit switches can be manipulated from the front and their positions read. The remaining programmer tubes are accommodated in pairs or groups of four in further boxes, together with their switches, and may be independently connected up to a decade. The switching tubes are housed in a separate box. The auxiliary circuits, being likewise made with printed wiring, can be similarly incorporated into the circuit.

The equipment described here has possibilities beyond that of controlling a coil-winding machine. Generally speaking, it can be employed whenever some switching action has to follow a specified number of occurrences, as long as these occurrences can be translated into light or voltage pulses. As examples we mention the control of machine tools by counting the number of revolutions of drive shafts, and the counting and sorting of objects on belt conveyors.

Switching can also take place after specified time intervals. The circuit is then fed with pulses from a pulse generator with a known and constant frequency. This possibility has been used for coating bulbs with an internal mirror. This is done by vaporizing a piece of aluminium within the evacuated bulb, after which the aluminium vapour settles as a deposit on the wall. The times required for evacuating the bulb, for heating the aluminium and for cooling down, depends upon the type of bulb treated. With the apparatus described these times can be rapidly preselected and changed. For small runs, especially, the flexibility of the apparatus again shows up to advantage over mechanical control.

Summary. In various incandescent lamps, e.g. those for projectors, the filaments consist of coiled sections linked by sections of uncoiled wire. Filaments of this kind can be made on a filamentwinding machine by decoupling the winding head from the motor drive for a few specified periods. Motor and coupling must be energized in accordance with a preselected programme, according to the type of filament required. The machine was hitherto mechanically controlled by cams. An electronic control has now been designed for the same purpose. It has the advantage that for small runs, e.g. in developing new types, the dimensions of the filament produced can easily be varied. The control is operated by voltage pulses derived from the periodic interruption of a light beam by a perforated disc fitted to the motor spindle. The pulses are applied to a decade counter circuit incorporating cold cathode tubes, type Z 70 U, as switching elements. Upon reaching a preset number (set by a switch in each decade), corresponding to a given length of coil (number of revolutions of the motor), the counter controls a transistorized switching unit, which switches the currents through motor and magnetic coupling. In this way the various stages of the winding programme are successively completed. The control apparatus is in principle also applicable to other manufacturing processes, e.g. in which the preselected numbers correspond to time intervals. As an example an installation for the internal silvering of bulbs is mentioned.

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 Laboratories, Eindhoven, Netherlands.

2698: P. B. Braun and J. L. Meijering: The copper-rich part of the copper-barium system (Rec. Trav. chim. Pays-Bas **78**, 71-74, 1959, No. 1).

The system Cu-Ba was examined up to 75% Ba by weight. At 675 °C there is a peritectic three-phase equilibrium Cu + liquid $\rightleftarrows$ $Cu_{13}Ba$. The compound $Cu_{13}Ba$ has KZn_{13} structure. At 550 °C there is a eutectic three-phase equilibrium liquid $\rightleftarrows$ $Cu_{13}Ba$ + Cu_xBa_y. Composition and structure of the latter could not be determined owing to experimental difficulties.

2699: W. Kwestroo: Spinel phase in the system MgO-Fe_2O_3-Al_2O_3 (J. inorg. nucl. Chem. **9**, 65-70, 1959, No. 1).

The three-component system MgO-Fe_2O_3-Al_2O_3 has been investigated at 1250 °C and at 1400 °C. The binary parts of this system, already described some twenty years ago, are reviewed and completed. The ternary diagram shows a broad spinel area: this area increases with increasing temperatures. The preparations do not contain Fe^{2+} ions, as was ascertained by analytical methods and by D.C. resistivity measurements. The resistivity has a fairly high value. The physical properties are roughly in agreement with the properties already found in systems investigated previously. Substituting Al^{3+} ions for Fe^{3+} ions in the spinel phase lowers the value of the magnetic saturation and the Curie temperature. An increasing amount of Fe_2O_3 in the spinel phase increases the magnetic saturation and the Curie temperature. Samples with the composition of a mineral called "hoegbomite" have been investigated and the possible structure of this mineral is discussed.

2700: J. Goorissen and F. Karstensen: Das Ziehen von Germanium-Einkristallen aus dem „schwimmenden Tiegel" (Z. Metallk. **50**, 46-50, 1959, No. 1). (Pulling germanium single crystals from a floating crucible; in German.)

Single crystals of germanium with a homogeneous impurity concentration can be prepared if enrichment or exhaustion of the impurity in the melt is prevented. This is made possible by means of the floating-crucible technique. In this technique a small crucible from which the crystal is pulled floats on the molten germanium in a larger outer crucible. Communication between both crucibles is made possible by means of a capillary allowing a continuous replacement of the solidified germanium. The theoretical yield of this technique is compared with that of the Czochralski method and of zone levelling. See also Philips tech. Rev. **21**, 193-195, 1959/60 (No. 7).

2701: H. J. G. Meyer: Theory of infrared absorption by conduction electrons in germanium (Phys. Chem. Solids **8**, 264-269, 1959).

It is shown that a theory of infrared absorption by conduction electrons which takes into account the structure of the conduction band and acoustical as well as optical intra-valley scattering can be developed without any serious approximation. In combination with an estimate of the possible influence of impurity scattering the theoretical results can be used for the determination of one of the acoustical and of the optical deformation potential constants. From available experimental data the approximate numerical value of the latter is determined. The general limits of validity of the theory are discussed.

2702: K. W. van Gelder: Fabricagebeheersing, I. Procesnauwkeurigheid en het stellen van toleranties (Sigma **5**, 15-19, 1959, No. 1). (Process control, I. Process accuracy and the specification of tolerances; in Dutch.)

First of two articles (see also **2719**) in which the author attempts to fuse the essentials of manufacturing processes (derived from analysis of such processes) with statistical concepts, and thus arrive at process control. This first article contains firstly a simple example of a process analysis (pressing and centreless grinding of plastic coil bobbins), which he uses to draw conclusions as regards the choice of tolerances. This is further considered in a second example. Finally, more complicated cases are dealt with in which so-called combined tolerances occur.

2703: G. H. Jonker: Analysis of the semiconducting properties of cobalt ferrite (Phys. Chem. Solids **9**, 165-175, 1959, No. 2).

From measurements of resistivity, activation energy, and Seebeck effect, an energy-level scheme is derived by which the semiconducting properties of $CoFe_2O_4$ can be described. These properties differ considerably from those of normal semiconductors, as the charge carriers are not free to move through the crystal lattice but jump from ion to ion.

2704: M. J. Sparnaay: On the additivity of London-Van der Waals forces (Physica **25**, 217-231, 1959, No. 3).

Calculations are given concerning the additivity of London-Van der Waals forces between two groups of atoms, the atoms being represented as isotropic harmonic oscillators. The results indicate that deviations of 10-30% from additivity can be obtained if only dipole-dipole interaction between oscillators of one group is assumed. The effect can be expected to be relatively large if the symmetry of the arrangements of the oscillators in the group is low, and it is dependent upon the relative spatial position of the groups.

2705: G. Diemer and P. Zalm: The role of exhaustion barriers in electroluminescent powders (Physica **25**, 232, 1959, No. 3).

Note concerning the apparent paradoxical effects of Mott-Schottky barriers in phosphor grains in an insulating medium. The explanation of the apparent paradox lies in the fact that electroluminescence occurs only in tiny localized spots where exhaustion barriers can indeed give rise to a local enhancement of the electric field.

VOLUME 21, 1959/60, No. 11 pp. 317-348 Published 27th September 1960

Philips Technical Review

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

A 75 cm RECEIVER FOR RADIO ASTRONOMY AND SOME OBSERVATIONAL RESULTS

by C. L. SEEGER *), F. L. H. M. STUMPERS and N. van HURCK.

523.164:522.6:621.396.722.029.63

The first half of the 17th century when, in rapid succession, the most outstanding discoveries in the stellar universe were made with the then newly invented telescope — the craters of the moon, the satellites of Jupiter, the phases of Venus, the stars of the galactic system, and so on — is looked on as the heroic era of astronomy. In centuries to come, the years around the middle of the 20th century may well be referred to in similar terms, for it was then that advances in radio engineering led to the construction of radio telescopes which, in their turn, brought to light unsuspected astronomical objects entirely different from those known hitherto.

Man's interest in the stellar universe dates from the earliest times. This interest led to the invention of telescopes and other optical aids with which visible radiation, and later infra-red radiation, could be observed more accurately and in greater detail. It was only about 15 years ago that astronomers realized that radio, too, could be a valuable means of widening our knowledge of the universe. In the Netherlands, for example, striking results have been obtained by radio observations at a frequency of 1420 Mc/s (wavelength 21 cm, a spectral line of hydrogen [1])). Radio observation is by no means limited to this frequency, however. On the contrary, the radio-frequency range usable for astronomical observations is very wide indeed. In the following we shall consider briefly some of the celestial objects thus brought within the purview of astronomers. We shall then describe a simple, stable and sensitive receiver for use in the frequency range from 200 to 500 Mc/s (150 to 60 cm) which has been used successfully for some years by the Netherlands Foundation for Radio Astronomy, at the Radio Observatory at Dwingeloo and elsewhere.

A principal requirement to be met by such a receiver is an extremely low inherent noise level. This is necessary in order to detect the extremely weak radiation received from most astronomical objects. The performance in this respect of the receiver referred to, though developed some years ago, ranks amongst the best that can be achieved in this field with conventional electron tubes. In latter years the parametric amplifier and the maser, too, have been developed as a means of reducing still further the inherent receiver noise contribution to the measured signal. In the coming years the parametric amplifier in particular will doubtless find wide application in radio astronomy. Its development, however, is too recent for any radio-astronomical experience to be yet available in the part of the frequency spectrum with which we are concerned.

Radio waves from outer space [2])

The spectral range covered by conventional optical astronomy comprises about two octaves

*) Netherlands Foundation for Radio Astronomy, Leyden Observatory. The work of this Foundation is made possible by financial support from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

1) See for example C. A. Muller, A receiver for the radio waves from interstellar hydrogen, I. The investigation of the hydrogen radiation; II. Design of the receiver, Philips tech. Rev. 17, 305-315 and 351-361, 1955/56. These articles may be referred to for a further explanation of various general facts and concepts of radio astronomy, such as the noise character of the radiation, the term antenna temperature, the directional diagram of a parabolic antenna, and so on.

2) A useful survey is given in the book: J. L. Pawsey and R. N. Bracewell, Radio Astronomy, Oxford Univ. Press, Oxford 1955. On equipment for radio astronomy see also: Proc. Inst. Radio Engrs., special issue, Jan. 1958.

(roughly 0.34 to 1.5 microns). The new frequency range which radio waves from outer space open up to astronomical investigations comprises some twelve octaves (*fig. 1*). At one end this range

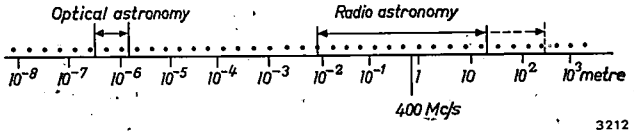


Fig. 1. Frequency spectrum of electromagnetic waves, indicating the two regions useful to the astronomer: the "optical" region (approx. 0.34 to 1.5 micron) and the radio region (approx. 8 mm to 20 m, possibly up to 300 m). The points mark divisions between octaves.

extends to millimetre wavelengths, where the radiation is absorbed by oxygen and water vapour. At the other end it extends to wavelengths of 15 or 20 metres. Longer waves are not generally passed by the ionosphere, although it has recently been shown that "windows" do occur now and then, allowing radiation of wavelengths longer than 100 metres to penetrate to the earth [3]).

Apart from the 21 cm radiation of hydrogen, radio astronomers have found no other spectral line in the radiation received from space. Theoretically, one might expect the analogous radiation from deuterium and from the OH radical, but investigations in this direction have produced no positive results. The detectable radiation possesses, rather, a continuous spectrum.

Understandably, man's thoughts turned first to the *sun* as a possible source of radio radiation. In 1894 Sir Oliver Lodge attempted to detect such radiation, but was not successful with the means available at that time. Much later, in 1942, Hey in England found that the sun was responsible for some originally unidentifiable interference in radar. Around the same time, Southworth in America succeeded in receiving radio waves from the sun. Since then, extensive observations have been made in this field, especially in England and Australia. We can now distinguish between the background radiation of the "quiescent" sun, and the intense radiation associated with sun spots and flares. It is even possible to follow in detail the movement of such a source of radiation through the solar atmosphere. Besides the radiation from the sun, thermal radiation from the moon and from the planets Venus, Mars, Jupiter and Saturn has also been observed in the radio spectrum in recent years; this is the very long-wave part of the Planck radiation emitted by these bodies, which possesses a maximum in the infra-red. In the case of Jupiter, additional radiation is emitted at certain times, perhaps as a result of thunderstorms far more violent than any known on the earth.

Against the more or less uniform background of the radio-wave radiation received from space, sources other than the sun were discovered. The radiation from these so-called point sources, i.e. points in the heavens emitting a much stronger radiation than their surroundings, was attributed to stars, which were at first called "radio stars". At that time, however, it was not possible to identify any of these objects with any optically observed star. The "radio stars" are much "brighter" than the sun, which plays such a prominent role in relation to the earth only because of its relatively short distance from it. Such stars, as for example Cygnus A presently to be discussed, may emit radio waves whose total intensity is 10^{36} times greater than from the sun.

With the more accurate determination of position made possible by means of interferometers (see below), it later proved possible to identify a number of sources with visible objects. For example, the supernovae — stars that flare up suddenly as a result of explosion — are powerful sources of radio-wave activity. One of these is the Crab Nebula (*fig. 2*), the outburst of which was observed by Chinese astronomers in 1054. The strongest source of radio waves, Cassiopeia A, is probably also an exploded star. (It has become the practice to denote a radio-wave source according to the constellation in which it is detected, and further by a roman letter in order of discovery.)

Another identifiable and often optically visible source of continuous radio-wave radiation is the ionized hydrogen gas in the plane of our Galatic system.

This radiation originates in so-called free-free transitions. Free electrons in uniform motion cause no radiation. When such an electron approaches close to a proton, it is accelerated and this gives rise to radiation. After deflection the influence of the proton again diminishes and the electron becomes free again. The theory of these free-free transitions has been worked out by Kramers.

Radiation is also observed from other galatic systems, a familiar example of which is the Andromeda Nebula. The strong source Cygnus A has been identified with an extra-Galactic object believed to be two galaxies in collision.

The mechanism giving rise to the radiation is far from known in all cases. The radiation received from

[3]) G. Reber and G. R. Ellis, J. geophys. Res. 61, 1, 1956.

Fig. 2. The Crab Nebula, the remains of a "supernova", the outburst of which was observed by Chinese astronomers in 1054 and which has now been identified as a source of intense radio waves. One millimetre of this photograph represents 2.3 seconds of arc. The picture was taken with the 100 inch telescope in California; blue filter (0.31-0.50 μ), exposure one hour.

the Crab Nebula was found, first by optical observations and later also at radio wavelengths, to be linearly polarized. It is therefore assumed that the radiation in this case is of the synchrotron type, i.e. radiation from electrons which travel at about the speed of light in helical orbits around the lines of force of a magnetic field. It is not yet known how the electrons acquire their energy in this mechanism or how the magnetic field is created.

In order to make a conjecture as to the cause of the radio waves from a particular source, it is desirable to compare the frequency-dependence of this radiation with that following from the theoretically assumed mechanism. It is therefore necessary to carry out measurements in several parts of the spectrum. Radio astronomers, who keep in close touch with each other, try to avoid duplication of effort and choose their part of the frequency spectrum with the object of supplementing as far as possible existing data. As a result, when the 25 m radio telescope at Dwingeloo in the Netherlands came into operation, a working frequency of 400 Mc/s was selected. This choice was also influenced, of course, by the possibility of building a good receiver for this range. Moreover, the relatively high frequency of 400 Mc/s was attractive in as much as the background radiation decreases more markedly with rising frequency than the radiation from some individual sources. At 400 Mc/s these sources therefore begin to show up more clearly against the background, and this effect is enhanced because the beam width of the antenna is smaller at such higher frequencies.

Antennae for receiving the radio waves

In the observations made with our receiver the following antennae were used (see *fig. 3*):

a) a parabolic radar reflector 7.5 metres in dia-

meter, which the Germans left behind in the Netherlands after the war; *b*) a horn antenna for calibration purposes (see below); *c*) a large parabolic

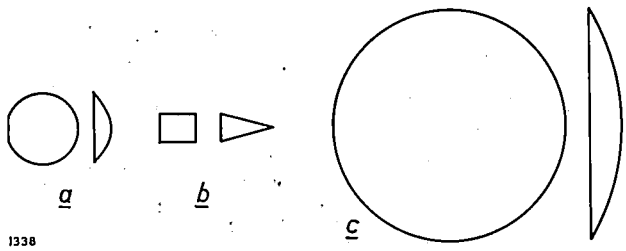


Fig. 3. Relative dimensions of the three antennae used in conjuction with the receiver described: *a*) Parabolic radar reflector of 7.5 m diameter. *b*) Horn antenna. *c*) Large, rotatable parabolic reflector of 25 m diameter, built at Dwingeloo in the years 1954 to 1956.

reflector, 25 metres in diameter, built at Dwingeloo in the years 1954 to 1956 [4]); see *fig. 4*.

The main lobe of the polar diagram for the radiation received by a parabolic reflector has an aperture angle that broadly depends on the wavelength λ and the diameter d of the reflector according to the formula:

$$\varphi = 70\ \lambda/d.$$

(φ in degrees, λ and d in the same units; this aperture angle is defined such that at the angle $\frac{1}{2}\varphi$ from the lobe axis the intensity has dropped to half.) A larger reflector will thus make it possible to distinguish proportionately more detail. The 25 m reflector at 75 cm wavelength (400 Mc/s) has a lobe width of only 2.1°.

This reflector, when it came into operation in 1956, was the largest rotatable radio telescope in the world; the only larger reflector is that of the Jodrell Bank radio telescope subsequently completed at Manchester, which has a diameter of 75 metres. The observations at Dwingeloo thus showed promise of providing a relatively very accurate picture of the radio sky.

In order to make even more accurate observations than are possible with large, rotatable reflectors, antenna systems are used that function as an interferometer. Ryle in Cambridge, and Mills and Christiansen in Australia have accumulated considerable experience with such systems. The Mills arrangement comprises two mutually perpendicular horizontal rows of dipoles (cross antennae). Systems of this kind have been made in which the rows of dipoles are 500 metres long; the beam width is 0.9° at a wavelength of 3 metres. Radio astronomers in the Netherlands, in cooperation with other Western European countries, are studying a plan for a much more ambitious system in which a length of several kilometres is envisaged. A drawback of such antennae is, of course, that they cannot be directed towards every point of the sky, so that one must wait for moments when interesting points pass through the narrow beam, which is movable only to a very limited extent.

The calibration of a large antenna, i.e. the derivation of the energy flux (flux density) from the received signal presents some difficulties [5]). Since, in the case of horn antennae the received signal is much better known — theoretically and experimentally — as a function of flux density, the above-mentioned horn antenna was used for the purpose of the absolute calibration discussed below.

The 75 cm receiver: general considerations

The design of a receiver is governed by the nature and strength of the signal to be received. The radiation which the astronomer wishes to receive has the character of *noise* such as is produced by thermal agitation of electrons in a resistance, or by the irregular emission of the cathode in a thermionic valve. Further, the radio astronomer must be able to deal with very weak signals. Whereas a field strength of e.g. 1 mV/m, or at least about 100 μV/m is required for a good broadcast signal, the astronomer is interested in detecting signals whose amplitude is 100,000 times smaller, corresponding to flux densities, integrated over the received spectrum, of the order of 10^{-20} W/m^2. The fact that the signals have the same character as the inherent noise of the receiver is an added difficulty. The inherent noise in a normal receiver may correspond to an input-resistance temperature of several hundreds or thousands of °K. Against this background it is required to detect signals corresponding to a temperature increment of about 1° due to a cosmic source. For this reason it is of paramount importance to reduce the inherent noise of the receiver to the minimum, and moreover to keep the gain of the

[4]) The radio telescope, Dwingeloo; six articles (in Dutch) by H. C. van de Hulst; B. G. Hooghoudt; R. J. Schor; W. Huisman; B. B. Schierbeek; G. H. Jöbsis; De Ingenieur **69**, O1-O19, 1957.

[5]) The flux density, expressed in watt/m^2 per c/s, is the energy incident per second and per unit bandwidth on a surface area of one square metre of the antenna. Where the source is extensive the concept "brightness" may also be introduced; this is the flux density divided by the solid angle subtended by the source at the point of observation.

Photo-montage by H. Kleibrink

Fig. 4. The 25 m parabolic reflector of the radio telescope at Dwingeloo.

receiver extremely constant whilst scanning the sky.

The following considerations indicate the most important fundamental limits to the reduction of the inherent noise level. After amplification, the signal is detected. The output power I of the detector fluctuates, even though the input noise

signal and the gain are constant. The magnitude of the fluctuation is given by

$$\overline{(I-I_0)^2}=\frac{pI_0^2}{\tau\Delta f}, \quad \ldots \quad (1)$$

where I_0 is the mean output power, τ the time constant and Δf the bandwidth of the receiver, and p a constant of the order of magnitude of unity, which depends to some extent on the frequency response of the receiver and on the detector characteristic. A signal, in order to be observed, must have an amplitude of at least the order of magnitude of these fluctuations.

The fluctuation, a measure of which is $\sqrt{\overline{(I-I_0)^2}}$, is evidently proportional to the noise level I_0. Further, the variation of a fluctuating quantity is inversely proportional to the root of the number of independent "readings" of this quantity; this accounts for the factor $\tau\Delta f$ in the denominator.

From eq. (1) it might be thought that, by increasing the time constant τ and the bandwidth Δf, one could arbitrarily reduce the lower limit of the measurable signals. The situation is not as simple as that, however. As the bandwidth increases the contribution of the inherent noise of the receiver increases more strongly than that of the detected signal, owing to the quadratic frequency-dependence of the noise factor (see below). Moreover, there is a greater chance of receiving an interference signal. Increasing the time constant also meets with difficulties. The time needed for a single measurement then becomes longer, and if the sky is continuously scanned the antenna must be made to move more slowly [6]). This is most undesirable for observations that extend over weeks or months, such as those involved in the accurate charting of regions of the sky. Again, with a larger time-constant value it is less easy to ascertain whether perhaps an interference peak has occurred that would vitiate the result of measurement. The choice of bandwidth and time constant must therefore be a compromise. The present receiver can have an optional bandwidth of 10 or 1 Mc/s, and the time constant can be varied, according to circumstances, from 0.1 to 10 sec. For some purposes the observation time can be lengthened by keeping the antenna directed for a certain time towards the same point and by averaging the recorded intensity. This has the same effect as a very large time constant, but one can now see from the recording whether an interference peak has occurred.

Before going into details of the receiver we shall discuss briefly the way in which the inherent noise of a receiver is determined.

[6]) Compare the entirely analogous problem of the continuous recording of an X-ray diffraction diagram: Philips tech. Rev. 17, p. 208 *et seq.*, 1955/56.

The noise factor and its measurement

The inherent noise which a receiver contributes to the noise level is expressed by the noise factor, F. This factor indicates by how many times the noise power at the output terminals is greater than that of of an ideal receiver (which itself contributes no noise) when no noise source is present other than the normal input impedance, at a standard temperature T_0. As this impedance we can take, for example, the characteristic impedance of the connection cable on which the receiver normally operates, namely 50 ohms. The noise power delivered by an impedance is proportional to its temperature T. We might also say, therefore, that owing to the noise contribution of the receiver the input impedance has an apparent temperature $F \times T_0$. The receiver noise thus gives rise to an apparent temperature increase of $(F-1)T_0$.

This leads directly to one of the methods of measuring the noise factor [7]). The output noise power P_1 is measured with the input impedance at a temperature T_1, and the measurement is repeated at a different temperature T_2, giving the noise power P_2. We may then write:

$$P_1 = a\{(T_1 + (F-1)T_0)\}, \quad \ldots \quad (2a)$$

$$P_2 = a\{(T_2 + (F-1)T_0)\}. \quad \ldots \quad (2b)$$

In accordance with the IRE Standards [8]) the standard temperature T_0 is taken to be 290 °K. a is a constant proportional to the receiver gain. We eliminate a from (2a) and (2b), and the result is:

$$F-1=\frac{P_1T_2-P_2T_1}{(P_2-P_1)T_0}. \quad \ldots \quad (3)$$

A practical procedure is first to take a measurement with a resistance of 50 ohms at room temperature connected at the receiver input, and then with a resistance whose value is 50 ohms at the temperature of liquid nitrogen (77.5 °K). The deflection of the detector is previously calibrated in terms of delivered power, by a method which will be touched on later.

Once the noise factor of the receiver has been determined, one can conversely derive from the measured output signal the temperature of the input

[7]) Various methods of determining the noise factor are discussed in N. van Hurck and F. L. H. M. Stumpers, An automatic noise-figure indicator, Philips tech. Rev. 18, 141-144, 1956/57.

[8]) IRE Standards on electron tubes: Definitions of terms, Proc. Inst. Radio Engrs. 45, 983-1009, 1957.

impedance. This applies not only to a true resistance but also to the radiation resistance of the antenna. The "antenna temperature" is, subject to certain conditions, equal to the temperature of the space upon which the antenna is directed, or, since the temperature of that space may differ from place to place, equal to the average temperature of the space seen by the antenna. (Each region is given a weighting factor corresponding to the size of its solid angle and its direction in the antenna polar diagram: the average temperature is thus a "weighted mean".) Frequently this is virtually the equivalent temperature of the radiation in the main direction of the antenna. The variation of this temperature over the sky, which is a measure of the radiation intensity, is what the radio-astronomer seeks to measure.

Details of the receiver

The receiver is essentially a normal superheterodyne type, having an intermediate frequency of 50 Mc/s. A silicon-crystal diode serves as the mixer stage. The block diagram of the whole receiver is shown in *fig. 5*. We shall now deal successively with some particular features, namely the high-frequency amplifier, measures to stabilize the amplification, and the detector.

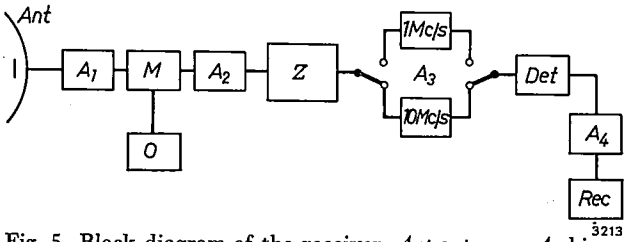


Fig. 5. Block diagram of the receiver. *Ant* antenna. A_1 high-frequency amplifier. *O* local oscillator. *M* mixer (silicon diode). A_2 intermediate-frequency preamplifier. *A* calibrated attenuator, adjustable in steps of 1 dB (see section on calibration). A_3 intermediate-frequency amplifiers with two optional bandwidths. *Det* detector. A_4 DC amplifier with variable time constant. *Rec* recording instrument.

The high-frequency amplifier

The most suitable circuit for the high-frequency amplifier is the cascode circuit [9]), which combines a low noise factor with high gain and great stability (no tendency to oscillation). In this circuit the cathode of an amplifying valve, whose grid is earthed, is connected for high frequencies to the anode of the preceding valve. For the frequency range in question the best valve in both places in the circuit is the type EC 57 disc-seal triode, which has the advantages of extremely short and constant electron-transit times and the high fraction of the saturation emission from an L-cathode that can be usefully employed [10]). The gain of two type EC 57 valves in cascode is about 36 dB at 400 Mc/s. A result of this high gain is that the following stage, the mixer, makes only a minor contribution to the noise factor.

The noise factor of a multi-stage amplifier is given by:

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \ldots + \frac{F_k - 1}{G_1 G_2 \ldots G_{k-1}} + \ldots, \quad (4)$$

where F_k is the noise factor of the k^{th} stage alone, and G_k the available power gain of that stage. It is assumed here that the F and G values do not depend significantly on frequency within the bandwidth of the receiver.

The high-frequency amplifier is shown in more detail in *fig. 6*. The antenna signal is applied via a 50 ohm coaxial cable to the amplifier input. Here an impedance transformer [11]) (T_1) enables the input impedance of the first valve (B_1), on which the noise factor is highly dependent, to be adjusted to the optimum value. The signal now arrives on the grid of the first valve, whose cathode is earthed for high frequencies. The cathode of the second valve is connected with the anode of the first. The stray capacitance of this point with respect to earth, which would form a short-circuit to the high-frequency anode current of the first valve, is effectively neutralized by tuning with a lecher system, T_2. (The two inner conductors in T_2 which are shown in fig. 6, together constitute for this purpose a single conductor, being interconnected at their ends. The arrangement in the same form as the impedance transformer T_1 is chosen here only because it facilitates the conduction of the necessary direct voltages to the valve electrodes — which must not of course be thereby short-circuited for high frequencies.) Similarly, the anode-grid capacitance of the first valve is neutralized by means of a simple coil, L, tunable with a sliding core. (The core is made of silver-plated bronze to limit eddy-current losses.) This is a variant of the familiar neutrodyne system, but its purpose is not to prevent instability due to feedback, the voltage gain of the first valve being too small (equal to unity) to produce such an effect.

[9]) Introduced and analysed by H. Wallman, A. B. MacNee and C. P. Gadsden, A low-noise amplifier, Proc. Inst. Radio Engrs. **36**, 700-708, 1948.

[10]) 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. The EC 57 valve used in our receiver has since been superseded by type EC 157, having a cathode of longer life. A general discussion of triode noise at high frequencies is given by G. Diemer and K. S. Knol, Philips tech. Rev. **14**, p. 236, 1952/53.

[11]) K. S. Knol and J. M. van Hofweegen, A universal adjustable transformer for U.H.F. work, Philips Res. Repts. **3**, 140-155, 1948.

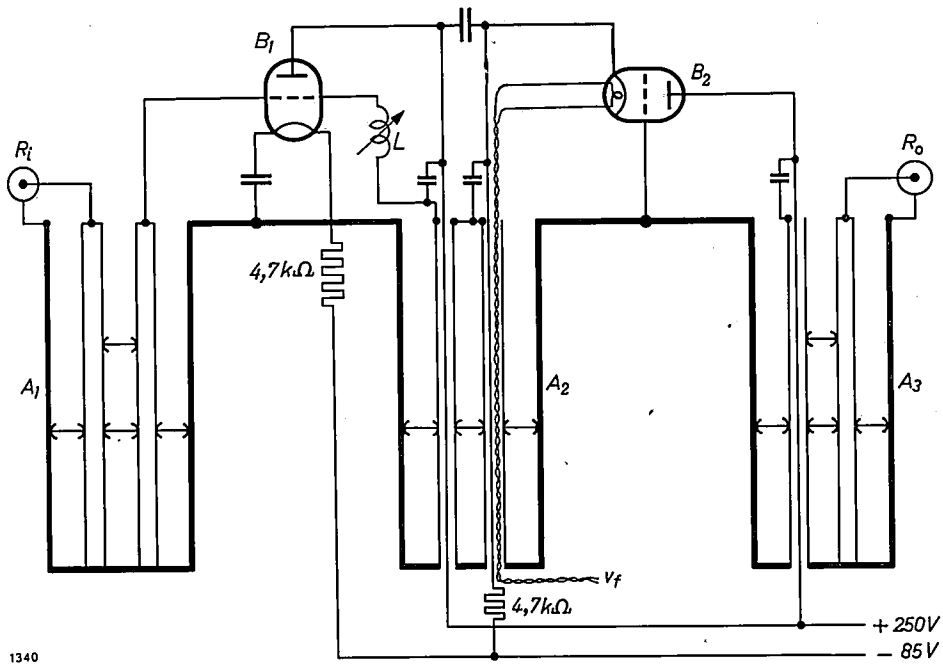


Fig. 6. Diagram of high-frequency amplifier. B_1 and B_2 two disc-seal triodes type EC 57. R_i coaxial cable of 50 ohms characteristic impedance, serving as input resistance. T_1 matching transformer in the form of a lecher system with two sliding shorting discs. T_2 lecher system for tuning the stray capacitance between the cathode of B_2 and earth; for convenience the same type is used as for T_1. L coil with plunger core for neutralization. T_3 matching transformer for a 50 ohm cable R_0, which functions as output resistance. The cathode of both valves receives a negative bias via a resistance of 4.7 kΩ; this bias determines the anode current (20 mA).

The object of the neutralization, in conjunction with the tuning of T_2, is rather to give the second valve a high input impedance, which, as experiments have shown, is needed to keep down the noise contribution of the valve. In the anode circuit of the second valve another impedance transformer (T_3 in fig. 6) ensures correct matching to the 50 ohm cable which conducts the signal to the mixer.

The arrangement of two triodes in cascode constitutes a system possessing the favourable properties of a pentode, but without the drawback of partition noise which attaches to the pentode. Whereas in the pentode the screen grid limits the coupling between input and output circuit (and thereby suppresses the tendency to oscillation) this is done in the cascode by the grid of the second triode, in which there is no current partition. Compared with a circuit having a single triode in the earthed grid configuration. the cascode arrangement possesses the advantage, typical of a pentode, of a much higher gain: the high cathode impedance which would be required for single triode to give a small noise factor, gives rise to negative feedback effects, which are of course not conducive to a high gain.

The noise factor found, using a selected EC 57 for the first valve, is plotted in *fig. 7* as a function of the tuning frequency. At 400 Mc/s the noise factor F was 2.5. Measurements on some thirty valves yielded an average value at this frequency of $F = 2.9$, with a spread of $\pm$ 0.3. The temperature $(F-1)T_0$, also shown in the figure, corresponds to the background, referred to above, against which an extra-terrestrial signal must stand out. The temperature corresponding to the above-mentioned noise factor $F = 2.5$ is $(F-1)T_0 = 435$ °K. With the aid of equation (1) we therefore find that, at a bandwidth of 10^7 c/s and a time constant of 10 sec, it is possible to detect an antenna temperature increment of $\sqrt{435^2/(10 \times 10^7)} \approx 0.05$ °K.

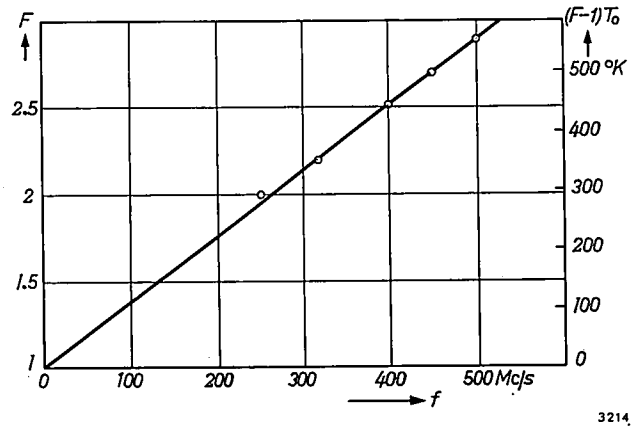


Fig. 7. Noise factor F of the receiver, measured on a selected EC 57 valve used as the first high-frequency amplifier (B_1 in fig. 6; this tube determines practically the entire inherent noise of the receiver, of which F is a measure). The graph is a plot of $F-1$ versus the tuned frequency f in Mc/s. At 400 Mc/s the value of F is 2.5. The points lie to a good approximation on the straight line defined by the equation: $F - 1 = 0.377 \times f/100$. On the right is a scale in terms of $(F-1)T_0$, i.e. the antenna temperature, equivalent to the inherent noise of the receiver.

The lower limit of the frequency range in which the receiver can operate is determined by the limited length (0.50 m) of the lecher systems used: tuning is not possible below about 200 Mc/s. The upper limit of the frequency range is determined by the neutralization: the simple method employed works well up to more than 500 Mc/s. With more elaborate devices it would probably be possible to reach 700 or 800 Mc/s.

Tuning with the lecher systems is so sharp that the incoming signal having the image frequency — which, due to frequency-changing in the mixing stage, also comes within the I.F. band — is attenuated by 45 dB with respect to the required signal.

The high-frequency amplifier is aligned by first setting all variables at maximum gain by means of a signal generator. Next, the first stage is adjusted for minimum noise factor by varying the impedance transformer T_1, the neutrodyne circuit and the tuning of the first anode (T_2). The automatic noise-figure indicator, described earlier in this journal [7]), is well-nigh indispensable for this purpose, since it shows immediately the effect of each change. Without it, a rather cumbersome procedure is necessary.

Special measures to ensure stability

We have seen that the amplifier must be very stable if spurious variations in the gain are not to be regarded as changes in the received noise power. Mains-voltage fluctuations are dealt with by a voltage stabilizer, which reduces a $\pm$ 10 % variation to a change of less than 0.1% in the supply voltage. The direct voltages for the valves obtained from this stabilized supply are held constant to within 1 in 10^5 with the aid of a stabilizer using a standard cell as reference. All amplifier stages are fitted with high cathode resistances, thus firmly establishing the bias of the valves, and also resulting in satisfactorily stable high-frequency characteristics. The detector used has the special feature that the level to be measured has no effect on the bandwidth and gain of the preceding amplifier valve (see below). Furthermore, the ambient temperature is kept as constant as possible in order to avoid temperature effects.

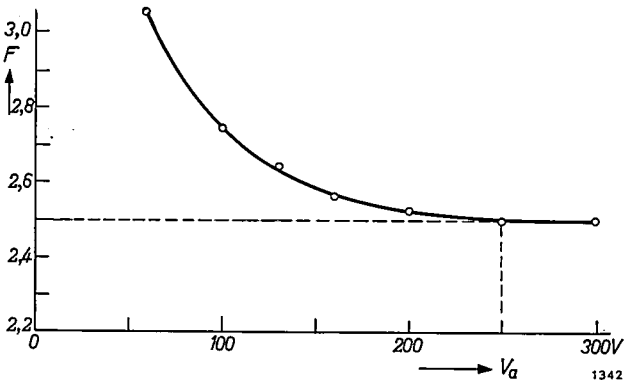


Fig. 8. Variation of the noise factor as a function of the direct anode voltage V_a on B_1. When V_a is varied, hardly any change is required in the adjustment of T_2 (see fig. 6). The value chosen was V_a = 250 V.

Special care is devoted to the aligning of the high-frequency amplifier. After adjustment for minimum noise factor, a test is made to ascertain to what extent the noise factor still depends on the biasing voltages. When the anode voltage of an EC 57 triode is gradually raised, the noise factor drops sharply between 60 and 160 volts; at higher voltages the effect is slight (*fig. 8*). An anode voltage of 250 V was considered sufficient for a favourable noise factor. Next, the variation of the noise factor as a function of the anode current was measured at 250 V (*fig. 9*). The minimum, which lies at 20 mA, was found to be fairly flat, which means that anode-current variations have little effect. It might now be asked whether, at this setting, the noise factor depends to any marked extent on the filament voltage. *Fig. 10* shows that this is not the case.

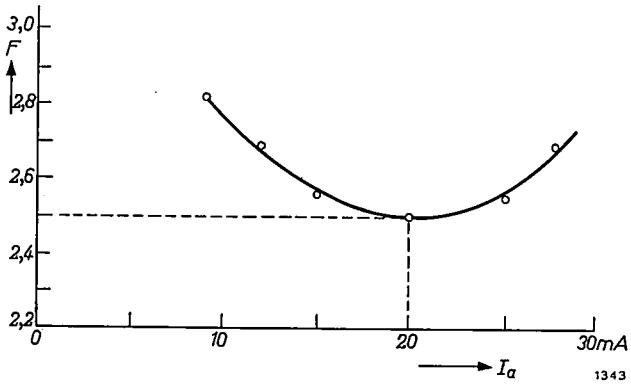


Fig. 9. Variation of the noise factor as a function of the anode current I_a. The optimum setting was found to be I_a = 20 mA; however, the minimum is rather flat.

It was mentioned above that the bandwidth of the receiver, which is normally 10 Mc/s, can be reduced to 1 Mc/s (viz. by switching-in an additional

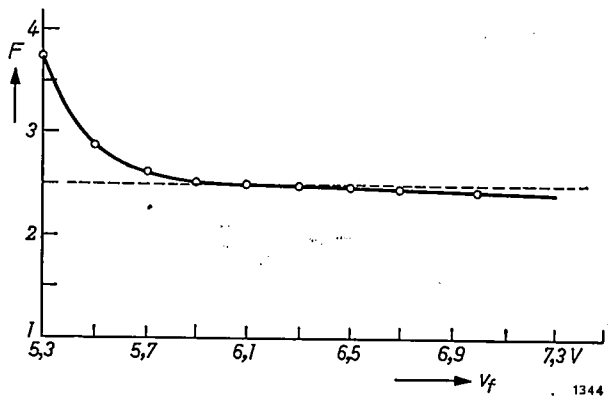


Fig. 10. Variation of noise factor as a function of valve heater voltage v_f.

band filter, see fig. 5) if interference makes it necessary. One might expect the noise factor at a bandwidth of 1 Mc/s to be appreciably smaller than at 10 Mc/s, considering that the noise factor rises quadratically with increasing (positive or negative) deviation from the tuning frequency [9]). However, the minimum to which the receiver is tuned is very flat, so that the noise factor variation for a frequency deviation of 5 Mc/s is of no significance.

Although the receiver is in fact suitable for the frequency range from 200 to 500 Mc/s, for the radio-astronomical observations to be described it has been operated solely at the fixed frequency of 400 Mc/s. This enables the frequency of the local oscillator to be stabilized with a quartz crystal. The crystal vibrates at a frequency of 29.17 Mc/s; this is trebled and twice doubled to obtain the required frequency of 350 Mc/s. For the latter frequency-doubling, where a fairly high power is demanded at the high frequency of 350 Mc/s, use is made of a small transmitting valve, type QQE 02/5.

The measures described resulted in a very stable receiver.

A known method of reducing the effect of gain fluctuations still further is to switch continuously between a reference noise source and the source to be measured (see [1]). Unless two receivers are used, however, this means losing half the observation time, This complication was therefore avoided here.

The detector

In the design of the receiver, it was endeavoured to maintain complete linearity up to the detector over a very wide range of signal amplitudes. Non-linear distortion of the noise signal, owing to the clipping of the peaks, for example, would cause errors in the intensity values. For this reason a constant-impedance type detector was chosen [12]) (*fig. 11*) so that, as opposed to normal diode detection, the last stage of the I.F. amplifier is not

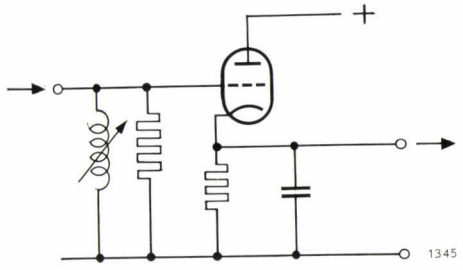


Fig. 11. Diagram of the detector (constant-impedance type). Owing to the presence of the *RC* network the anode current is a measure of the average amplitude over several cycles of the signal voltage applied to the grid (anode detection). Since hardly any grid current flows, the preceding stage is not loaded and the gain linearity is maintained up to peak voltages of e.g. 30 V.

[12]) W. K. Squires and R. A. Goundry, Electronics **25**, April 1952, p. 109.

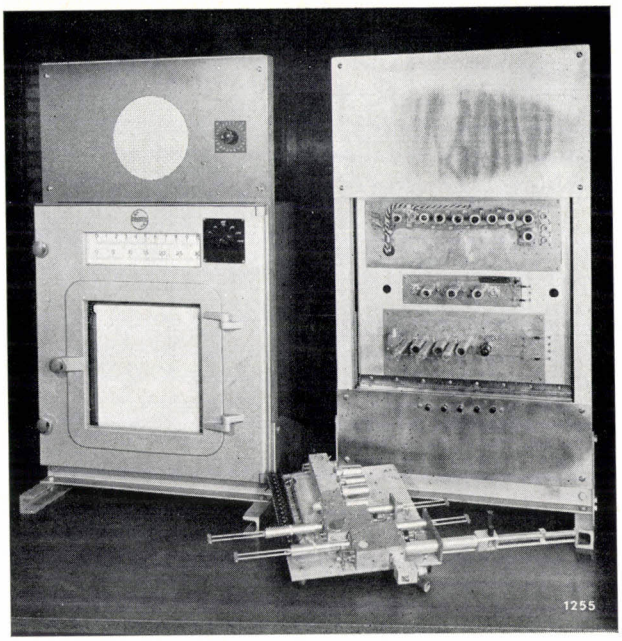


Fig. 12. The complete receiver. The chassis in the foreground contains the high-frequency amplifier (somewhat differently constructed but essentially the same as described in the text), the mixer diode and the IF preamplifier. This assembly is mounted at the pole of the parabolic reflector (fig. 4), where the mast is fixed, the other end of which carries, in the focal point of the reflector, the antenna proper (a kind of horn) and through which the coaxial connecting cable passes. The panel on the right contains the power-supplies, the detector and D.C. amplifier and the two I.F. amplifiers (for 10 and 1 Mc/s, respectively), On the left, the recording instrument and the monitoring loudspeaker.

subjected to a varying load when the signal amplitude varies. For the same reason, instead of the 1.5 W valve conventionally used in wide-band amplifiers, a type E 81 L valve is fitted in the last I.F. stage, which has a power of 6.5 watt and delivers a sinusoidal output voltage up to 30 V without significant distortion. True, the r.m.s. value of the signal applied to the detector amounts to no more than 3 V from the most powerful astronomical objects. Owing to the noise character of the signal, however, much higher instantaneous values appear at the output of the I.F. amplifier: the amplitude 12 V occurs, for example, during 0.01% of the time.

The detector voltage is fed to the recording meter via a DC amplifier, which is stabilized by negative feedback and by periodic switching to a reference signal [13]). This amplifier serves not only to produce a low output resistance, for matching to the recording instrument, but also enables the zero point of the scale to be placed far outside the reading range of the meter or strip recorder: the signal variations can thus be recorded enlarged up to 100 times. Furthermore, it is a simple matter to make the time constant of the DC amplifier variable. In this way the time

[13]) S. Landsberg, A general-purpose drift-free D.C. amplifier, Philips Res. Repts. **11**, 161-171, 1956.

constant of our receiver was made variable, as mentioned above, between 0.1 and 10 sec.

A photograph of the complete receiver appears in *fig. 12*.

Calibration of the receiver

The calibration of the receiver comprises a number of somewhat diverse procedures: measuring the noise factor of the receiver, determining the detector "law" (i.e. the relation between the output voltage of the detector and the noise power applied as signal to the receiver input) and calibrating the antenna.

Measuring the noise factor

The principle of measuring the noise factor having been described above, we shall deal here at greater length with a single, and rather unusual, detail of the set-up. For the measurement it is necessary to connect to the input a resistance of 50 Ω, which must first be held at room temperature and subsequently at the temperature of liquid nitrogen. A pure resistance, without inductance or capacitance, at a frequency of 400 Mc/s is feasible only as a coaxial resistance. However, not every coaxial resistance can tolerate immersion in liquid nitrogen. After each cycle of cooling and re-heating, such resistances do not usually return to exactly the same resistance value. A coaxial resistance that gives reproducible results under these conditions was constructed as follows. A ceramic rod is coated at one end with a thin layer of metal of length about 15 mm. This has a resistance of 50 Ω. The metal layer is a mixture of gold and platinum, prepared from a suspension marketed under the name "Zilverglans" (Regout, Maastricht). This preparation is applied with a brush and subsequently baked-in. The rest of the ceramic rod is then given a coating of silver making contact with the metallized end. It now constitutes the inner conductor of a coaxial system, the rest of which is formed by the coaxial cable to the receiver. The outer conductor is formed by a copper tube, dimensioned so as to obtain a characteristic impedance of 50 Ω. At the position of the resistance layer the outer conductor is tapered (*fig. 13*) and at the end, at *a*, the inner and outer conductors are soldered together. The tapering provides at each point, as it were, a match to the still remaining portion of the ohmic resistance.

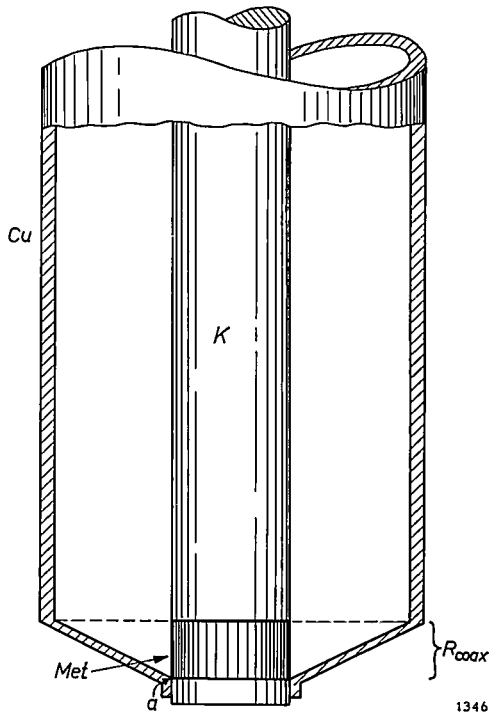


Fig. 13. Construction of the coaxial resistance, R_{coax}, designed for immersion in liquid nitrogen. *Cu* copper outer conductor. *K* silver-plated ceramic rod, coated at one end with a special metal layer *Met*.

The equivalence of the cooled resistance with a similar coaxial resistance held at room temperature is checked with an admittance bridge. If a small difference exists, the resistance at room temperature can be made exactly equal to the other by means of an impedance transformer [11]), like that described above under the sub-heading *High-frequency amplifier*. (This transformation is applied to the resistance at room temperature because losses occurring during this process must be regarded as contributions to the total resistance, due to resistances at room temperature.)

The detector law

For determining the detector law — which must be known in order to be able to measure the noise factor — a calibrated attenuator is incorporated in the I.F. amplifier; see fig. 5. This is a wide-band attenuator, with a 50 Ω termination at both ends. It contains five stages, giving an attenuation of 1-2-3-4-10 dB, respectively, and is operated by push-buttons; in this way an attenuation up to 20 dB can be introduced in steps of 1 dB. The attenuator was tested at the Naval Research Laboratory in Washington, D.C.; the deviation from the nominal attenuation was found in no case to be greater than 0.1 dB.

With no input voltage, i.e. solely as a result of the inherent noise of the receiver, the detector at a particular setting of the apparatus delivered a D.C. output of 0.966 V, with the attenuator set at 0 dB. If we now switch in 1 dB attenuation and apply a noise signal of sufficient strength to produce the same meter deflection, we find that, on returning to an attenuation of 0 dB, the D.C. output is 1.173. Proceeding along these lines we obtain the data presented in *Table I*.

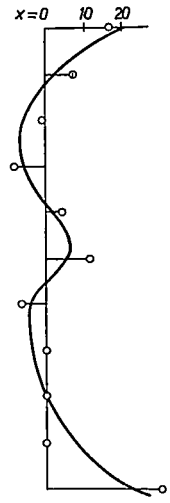
The steps of 1 dB are obviously much too coarse for radio-astronomical observations. We can interpolate with the following formula derived from the table:

$$1000\,E = 950 + 202\,a + 13\,a^2 + a^3 \ldots \quad (5)$$

The individual values of E in the table show small irregular deviations x from this formula. These are attributable to reading errors in E or to errors in the attenuator (errors as small as 0.03 dB are sufficient

Table I. The detector law determined by means of the calibrated attenuator: direct voltage output E as a function of input noise power a.

Applied input noise power a dB	Equiv. temperature of input resistance °K	D.C. voltage E at output V	Deviation x from eq. (5) mV	Graphic interpolation of x
0	290	0.966	+ 16	
1	365	1.173	+ 7	
2	460	1.413	— 1	
3	579	1.692	— 8	
4	728	2.034	+ 4	
5	917	2.421	+ 11	
6	1155	2.84	— 6	
7	1452	3.34	0	
8	1830	3.91	0	
9	2310	4.55	0	
10	2900	5.30	+ 30	



to explain the discrepancies). In practice a correction is applied to the values interpolated in accordance with (5); this correction is obtained by graphic interpolation of the x value (see graph in the table).

Calibration of the antenna

With the aid of the detector law the measured radiation intensities are expressed in relation to a fixed datum level, the inherent noise of the receiver. To determine the absolute radiation intensities, the primary unknown factor being the radiation-receiving power of the antenna, it is sufficient to know accurately the absolute value of the flux density of a single radio star. This antenna "calibration" is necessary in order to be able to compare the results of different radio observatories, and particularly in order to combine measurements at different wavelengths into a reliable spectral distribution. As mentioned in the introduction, this is important for the purpose of testing theoretical predictions regarding the mechanism of the radiation.

The "calibration star" chosen was the radio-source Cassiopeia A already referred to. Apart from the short waves, where the sun has the greatest intensity, this star is the most powerful source of extra-terrestrial radio waves in the whole radio spectrum.

The calibration was carried out in several steps, using the horn antenna mentioned earlier. The dimensions of this antenna were: axial length 5.625 m (= 7.500 λ), vertical aperture 3.028 m (= 4.038 λ) and horizontal aperture 3.756 m (= 5.008 λ). From these dimensions a gain of 20.95 dB is calculated with respect to the isotropic antenna [14]). The gain factor of the 7.5 m reflector was then determined by taking alternate measurements on a single source (viz. the sun) with this reflector and with the horn antenna; the gain found was 27.1 dB. The flux density of Cassiopeia A was now measured with a simple interferometer arrangement, using both antennae placed at a distance of 20 λ apart. With this arrangement the beam of the radiation received is so narrow as to neutralize the more continuously distributed background radiation, and the gain factor of the whole is known from the preceding steps. The result obtained was a flux density of 56.2×10^{-24} W/m² per c/s, corresponding to an antenna temperature of about 800 °K. (A detailed description of this measurement is given in the article in B.A.N. No. 472, referred to under [15])).

During the operation of the receiver installation the antenna is directed towards Cassiopeia A at regular intervals of, say, an hour, to check the amplification of the receiver. The receiver has proved to be so stable that no adjustment is required for weeks on end.

Some radio-astronomical observations made with the receiver described

As mentioned in the introduction, the Netherlands Foundation for Radio Astronomy have had the receiver described in operation for some years, mostly at Dwingeloo. During this time the receiver has given no difficulties whatsoever. Some results obtained in conjunction with the Observatory of Leiden University [15]) are briefly mentioned below.

An observation was made with the aid of the 7.5 m reflector of the partial eclipse of the sun on 30th June 1954. Owing to severe interference from a neighbouring transmitter the bandwidth had to be

[14]) S. A. Schelkunoff and H. T. Friis, Antenna theory and practice, Wiley, New York 1952.

[15]) C. L. Seeger, 400 Mc/s partial eclipse observations on 30 June 1954, Bull. Astron. Inst. Netherlands (B.A.N.) No. 461, 3 Dec. 1955, 273-283.
C. L. Seeger, G. Westerhout and H. C. van de Hulst, The flux densities of some radio sources at 400 Mc/s, B.A.N. No. 472, 26 Nov. 1956, 89-99.
C. L. Seeger, A tentative measure of the flux density of Cassiopeia A at 400 Mc/s, B.A.N. No. 472, 26 Nov. 1956, 100-104.
C. L. Seeger and G. Westerhout, Observations of occultations of the Crab Nebula by the moon at 400 Mc/s, B.A.N. No. 478, 9 Sept. 1957, 312-316.
C. L. Seeger, G. Westerhout, T. Hoekema and R. G. Conway, shortly to be published in B.A.N.

limited to 1 Mc/s (to be more exact: 0.9 Mc/s) and a time constant of only 0.1 sec was used in the detector in order to distinguish more clearly the effect of disturbances, originating for example from car ignition systems. Observation began two hours before the eclipse and was continued for one hour afterwards. The reflector did not follow the sun automatically, but was adjusted at two-minute intervals. *Fig. 14* shows a comparison of the observed eclipse with that calculated geometrically. A smooth curve can be drawn through the measured points with an error within one scale unit, except at the beginning and the end, where the error may be $1\frac{1}{2}$ scale units. During this eclipse the sun was very calm. The average radio-brightness of the sun at 400 Mc/s, calculated for a diameter of about 30 minutes of arc, corresponded to that of a black body having a temperature of 6×10^5 °K. The reference level for this measurement was a resistance at room temperature, which was switched-in at least once every ten minutes.

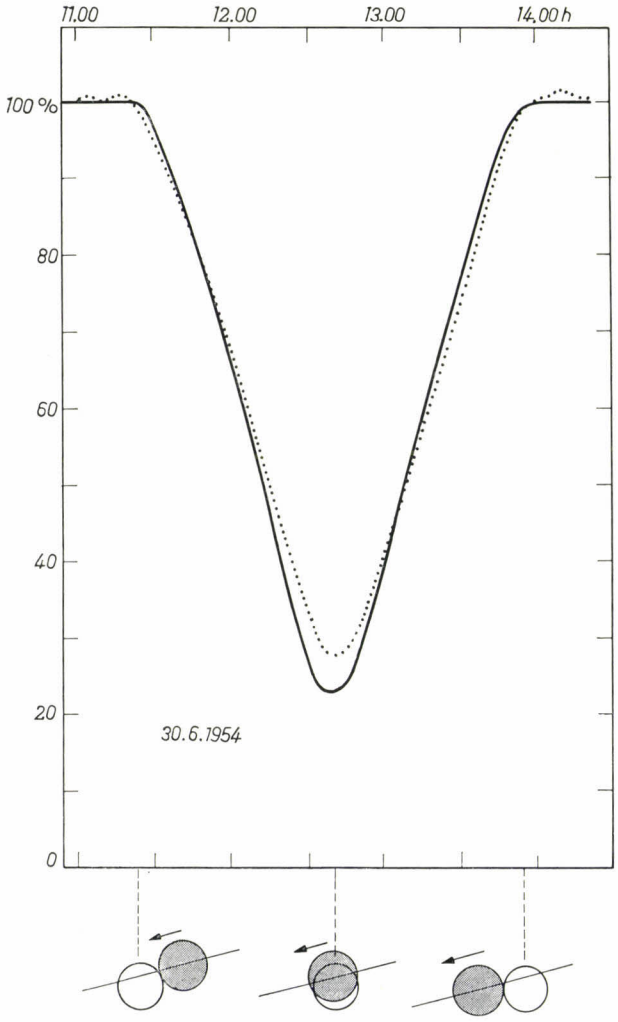


Fig. 14. Relative change of solar radiation at 400 Mc/s (arbitrary units), measured during the partial eclipse of the sun on 30th June 1954 at Waalsdorp (dotted curve). The solid curve is the geometrically calculated curve of the eclipse (percentage size of the visible surface of the sun). The position of the moon in relation to the sun at the beginning, zenith and end of the eclipse are illustrated below. Comparison of the measured and calculated radiation intensities shows that the radio "brightness" of the sun is not entirely constant over its surface (see the first article quoted under footnote [15]).

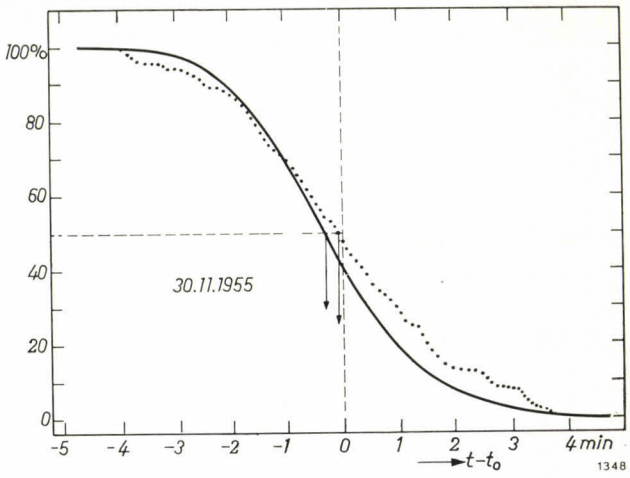


Fig. 15. Intensity of the 400 Mc/s radiation from the radio source *Taurus A* (the Crab Nebula, see fig. 2) recorded at Dwingeloo during the first part of the occultation by the moon on 30th November 1955 (dotted curve). The solid curve is the optically observed curve of the eclipse. t_0 is the calculated time at which the middle of the twin star, near the centre of the Nebula, disappeared behind the moon. The observation showed that the radio centre of the Crab Nebula differs only by about 12 seconds of arc from the optical centre point (in green light).

Another event observed with this receiver was the occultation of the Crab Nebula by the moon on 3rd and 30th November 1955. The large radio telescope at Dwingeloo was not entirely ready at that time and was unable to follow the occultation automatically. Since the event would not recur in our latitude for many years, the construction work was interrupted for four days in order to make the observation. The recording (*fig. 15*) showed that the radio centre of the Crab Nebula (Taurus A) virtually coincides with the optical centre. Broadly speaking, the angular extent of the Nebula at 400 Mc/s is the same as that observed optically, but the angular distribution of the radiation is much flatter than the distribution of (green) light. The observation of the eclipse also showed that no detectable refraction of the radiowaves is caused by any ionized layer in the moon's atmosphere. Earlier calculations had indicated that such refraction was unlikely.

Large parts of the sky have been charted for 400 Mc/s radiation. Each time the antenna was swung through 30° azimuth at a speed of about 10° per minute. After each such scan the elevation was increased 1°. *Fig. 16* shows a recording obtained with a single scan of the antenna (in this case under slightly different conditions of movement). The chart constructed from all these observations, which extended over several months, is reproduced as *fig. 17* on pages 332-333, together with a detailed

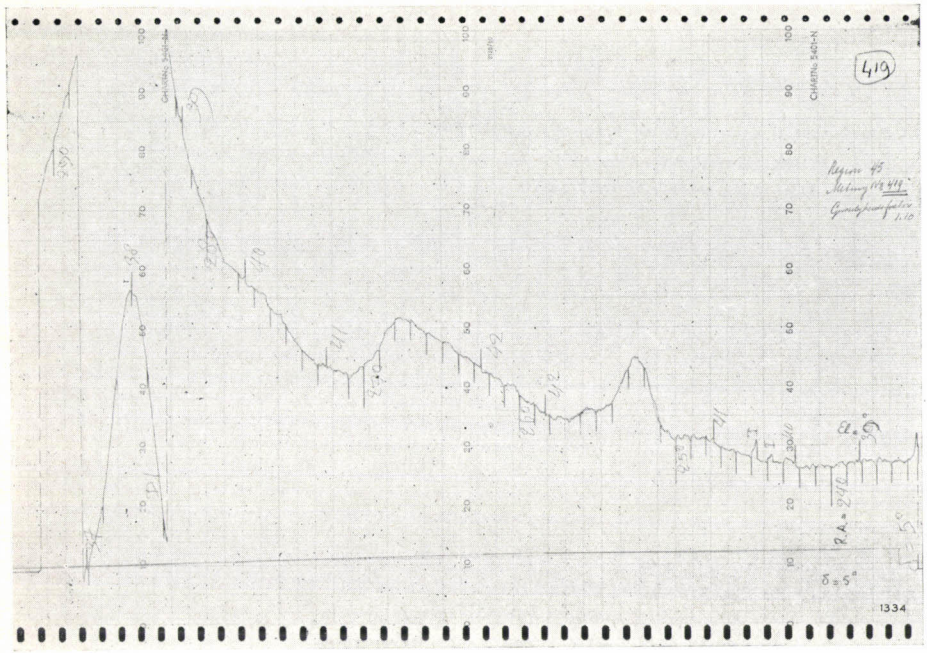


Fig. 16. Recording obtained with the receiver described for a single scan with the 25 m radio telescope at Dwingeloo. The path of the antenna beam for this recording is shown as a dotted line on the overlay sheet for fig. 17. In this case the angle of declination was kept constant at $\delta = 5°$, and the antenna moved at a speed of 8° in right ascension. The degrees marked below the recorded curve refer to right ascension, and those above refer to elevation. In order to be able to record the intensity peak that occurs as the antenna sweeps through the Milky Way, the scale is temporarily displaced by 100 scale divisions.

description. Owing to the very high sensitivity of the installation (low noise factor and high radiation sensitivity) and to the high resolving power of the 25 m reflector, the results now achieved are very much better than was possible ten years ago. The new chart, which incidentally by no means contains all the detailed information extracted from the observations (see below), may be compared with the results of Reber [16]) in 1948; see *fig. 18*.

The chart, fig. 17, is drawn in the form of contours of constant received power (isophots). This power, as already mentioned, is a weighted mean over the antenna beam; it further represents a mean over the frequency band of the receiver, although within the bandwidth of 10 Mc/s the variations of most of the sources observed at 400 Mc/s are negligible. The directional diagram of the antenna contains side lobes in addition to the central beam (see [1]) and [14])). It is possible to compensate for their influence but this correction has not been made in the chart reproduced. As in optical observations, twinkling due to atmospheric influences is also found in radio observations, particularly from the northern direction, where aurora effects occur. This direction was therefore avoided as far as possible; little trouble is experienced from twinkling when the antenna is directed towards the south.

The numbers on the isophots are proportional to the received power; the unit is a flux density of 9.1×10^{-26} W/m^2 per c/s, or an antenna temperature of 1.3 °K. To make the chart easier to read, the isophots are drawn with fairly large intervals. In this way the chart gives a clear picture of the broad structure of the 400 Mc/s radiation received from space, but many small details derived from the observations are lacking in this reproduction and only the strongest "point sources" are distinctly recognizable. The legend to the chart explains various particulars to be read from the contours. The "tails" of the general radiation from our galactic

[16]) G. Reber, Cosmic static, Proc. Inst. Radio Engrs. **36**, 1215-1218, 1948.

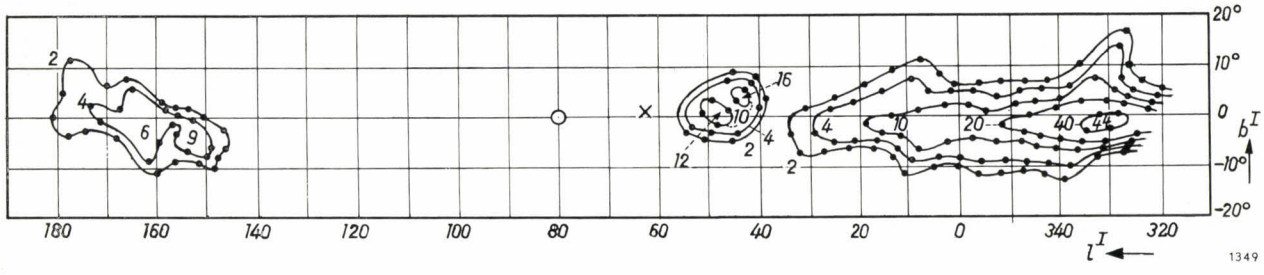
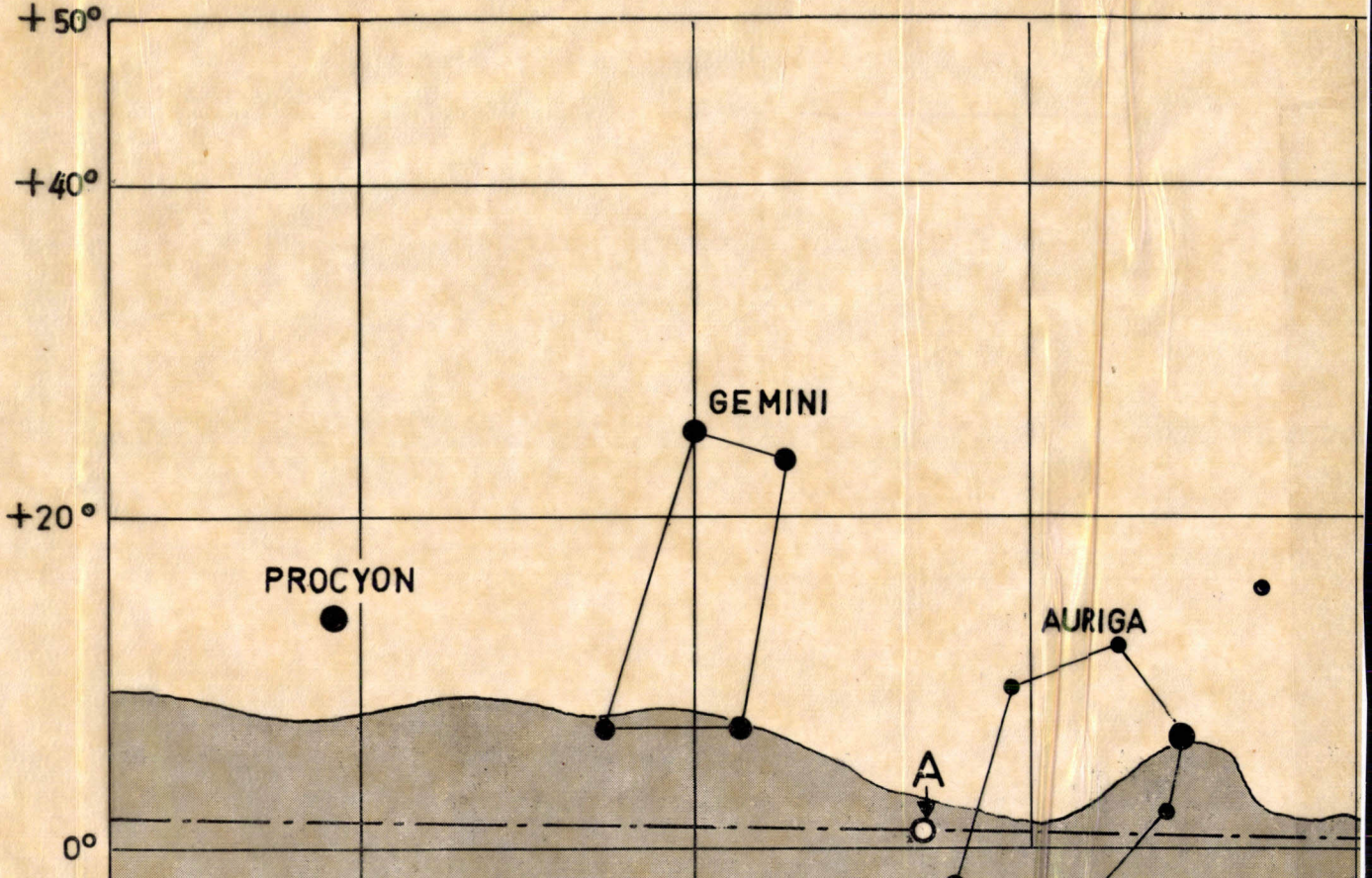
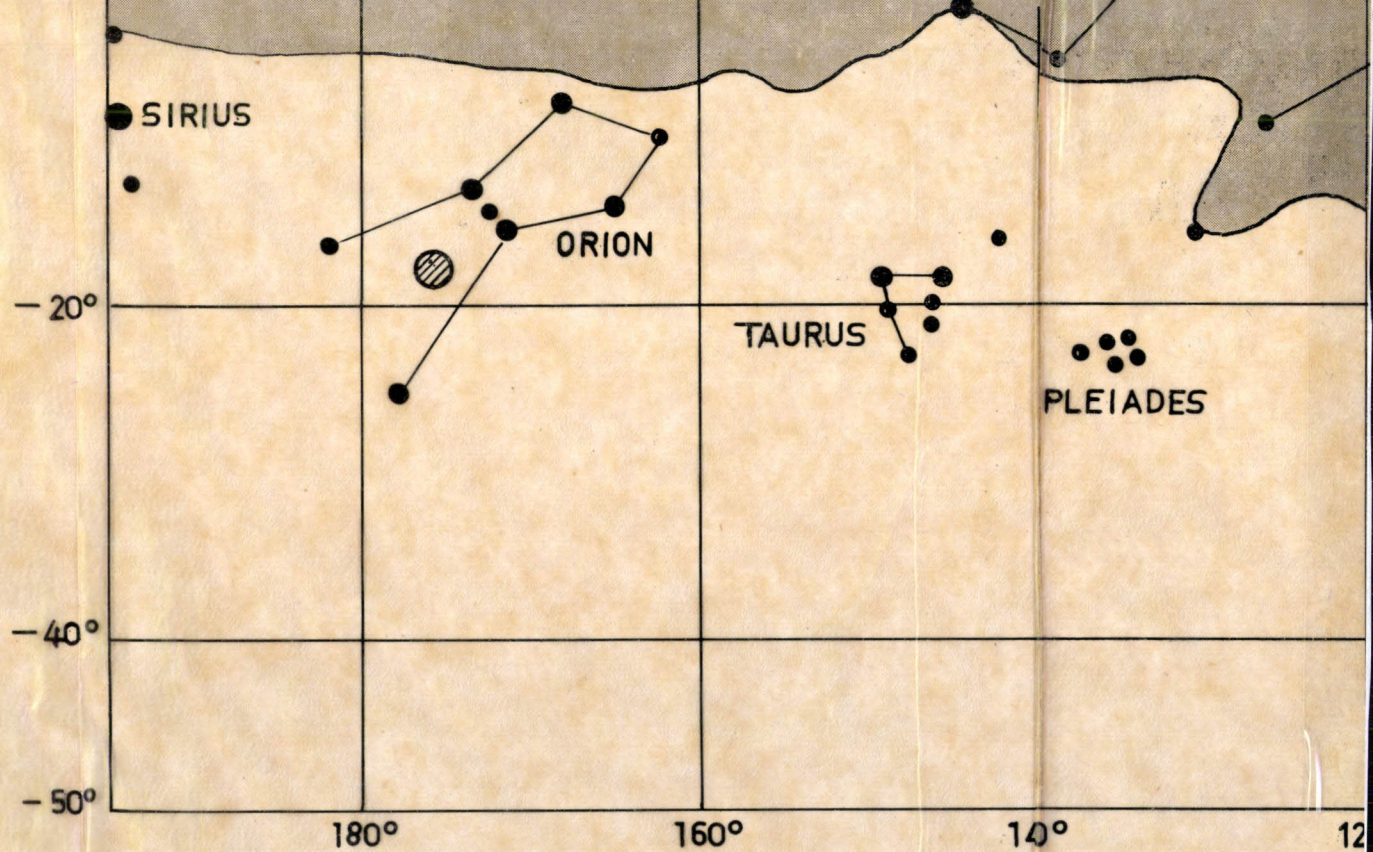
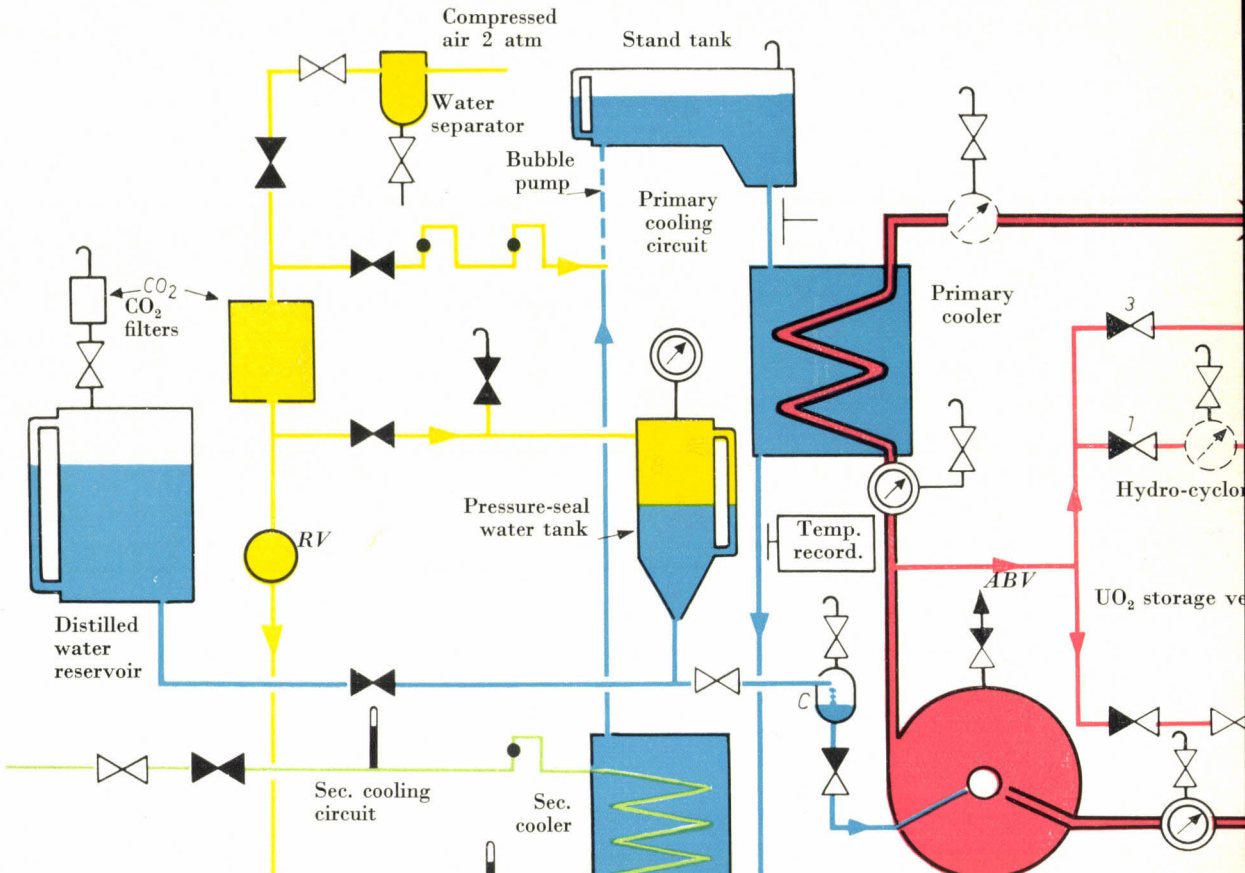
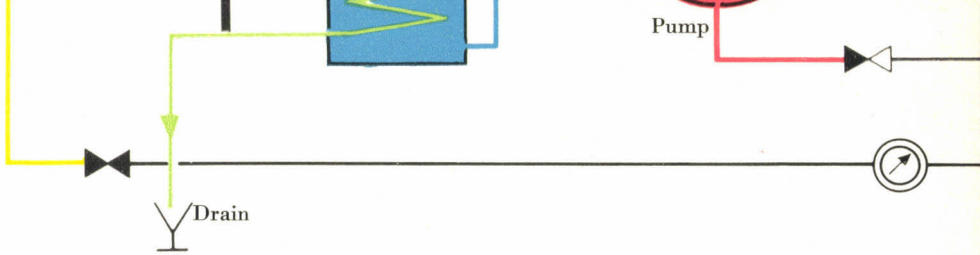


Fig. 18. Chart of the sky for 480 Mc/s radiation, constructed from observations by Reber [16]) in 1948. The chart shows the part of the sky between the same galactic longitudes l^I as in the new chart in fig. 17. The progress illustrated by the latter figure, as a result of the much greater resolution and higher sensitivity of the equipment at Dwingeloo, is evident.


+50°
+40°
+20°
0°
GEMINI
PROCYON
AURIGA
A


SIRIUS
ORION
TAURUS
PLEIADES
−20°
−40°
−50°
180°
160°
140°
12


Compressed
air 2 atm
Stand tank
Water
separator
Bubble
pump
Primary
cooling
circuit
CO_2
CO_2
filters
Primary
cooler
3
7
Hydro-cyclon
RV
Pressure-seal
water tank
Temp.
record.
ABV
UO_2 storage ve
Distilled
water
reservoir
C
Sec. cooling
circuit
Sec.
cooler



Legend:

Pipes outlined in black belong to the reactor circuit; the others are ancillary lines for control, cooling, etc.

Red: UO_2-water suspension Green: main water
Blue: distilled water Yellow: compressed air

Valve, not on control panel

Valve operated by remote control

Valve, on control panel

Resistance thermometer

Mercury thermometer

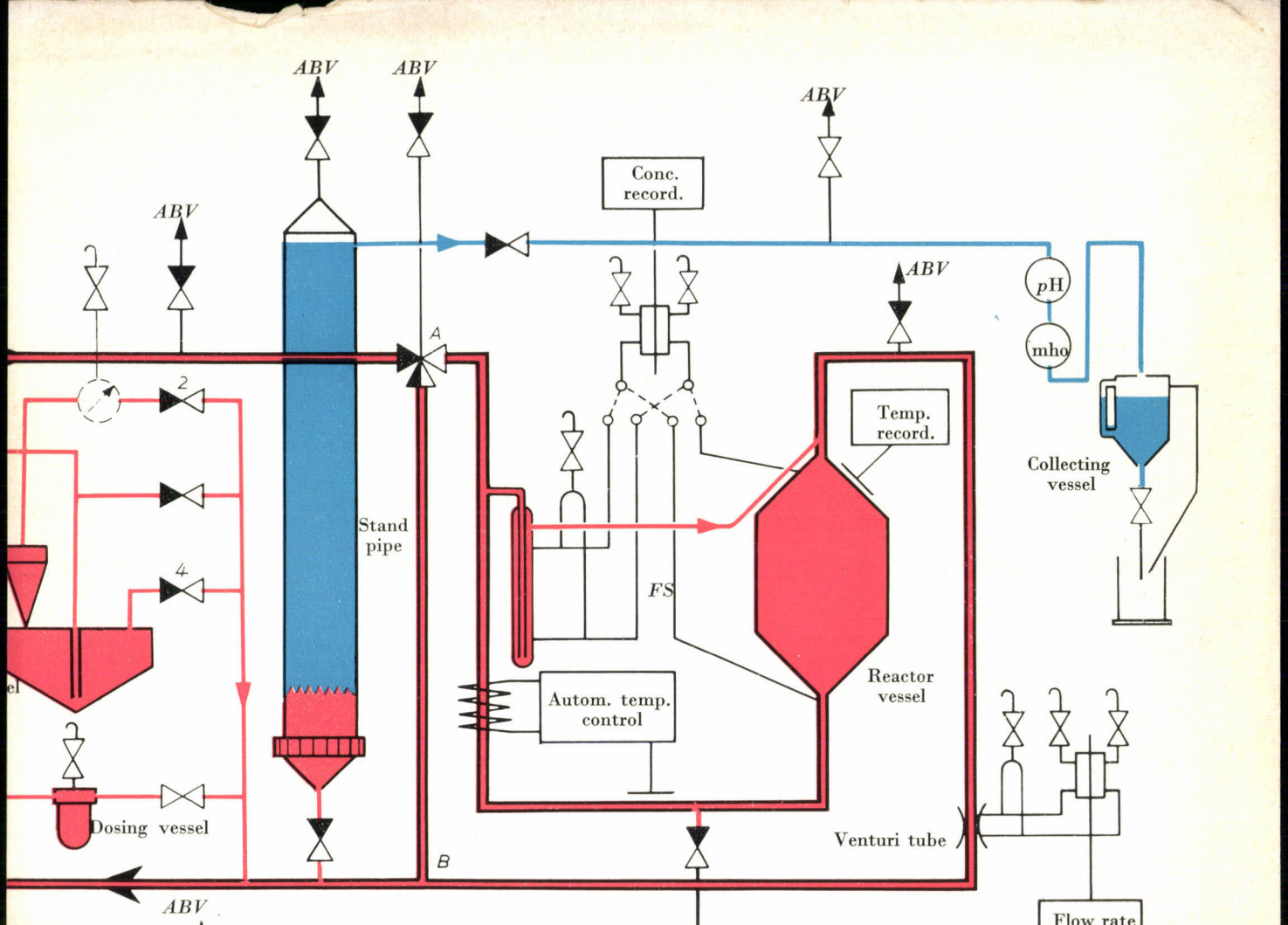
Manometer, *in situ* on reactor and also on control panel

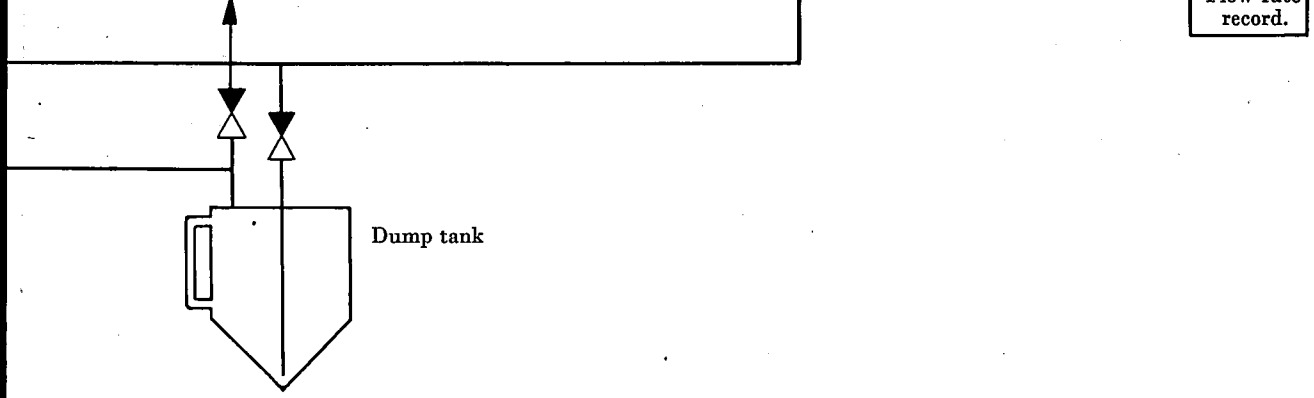
Manometer, only on control panel

Manometer, *in situ* on reactor only

Float-type flowmeter

ABV


ABV
ABV
ABV
ABV
ABV
Conc. record.
pH
mho
Collecting vessel
Temp. record.
A
2
4
Stand pipe
FS
Reactor vessel
Autom. temp. control
Dosing vessel
Venturi tube
B
ABV
Flow rate



ectric differential pressure pick-up

r bleeder

r bleeder via bleeder vessel, into
hich no CO_2 can enter from outside

evel gauge

Fig. 21. Flow sheet of the subcritical suspension reactor installed in one of the laboratories of N.V. KEMA at Arnhem. Through the reactor circuit (see legend) a suspension of UO_2 particles in distilled CO_2-free water is kept in continuous circulation. The pump speed is not variable; the flow rate in the reactor vessel is varied by means of a 3-way valve *A* which controls the fraction of the circulated fluid flowing through the shunt pipe *AB*. The flow through the reactor vessel is measured by a venturi tube (with an electric differential pressure pick-up) and continuously recorded. Also recorded are the temperature in the reactor vessel, the temperature of the cooling water upon leaving the primary cooler and the average concentration of the suspension in the vessel. The very sensitive and stable differential pressure pick-up used for the latter purpose can be connected either to the flow simulator *FS* or to the reactor itself, as required. The automatic temperature control is upstream of the reactor vessel. By opening valves *1* and *2* and setting the hydro-cyclone in operation, a very large proportion of the UO_2 particles can be removed from the fluid and collected in the UO_2 storage vessel. The latter can be emptied by closing valves *1* and *2* and opening *3* and *4*. By means of the dosing vessel (like the UO_2 storage vessel, detachable) small known amounts of UO_2 can be added to the circuit. The flow of pressure-seal water is measured with a droplet flowmeter *C*. The acidity (*p*H) and the electrical conductivity (mho) of the water are measured as it leaves the stand pipe. If necessary, the reactor circuit can be completely drained, the UO_2 particles being first transferred to the UO_2 storage vessel and the water then drained into the dump tank. This tank can be emptied by compressed air taken via a reducing valve *RV* from the air line which also serves the bubble pump in the primary coolant circuit and the pressure-seal water line.

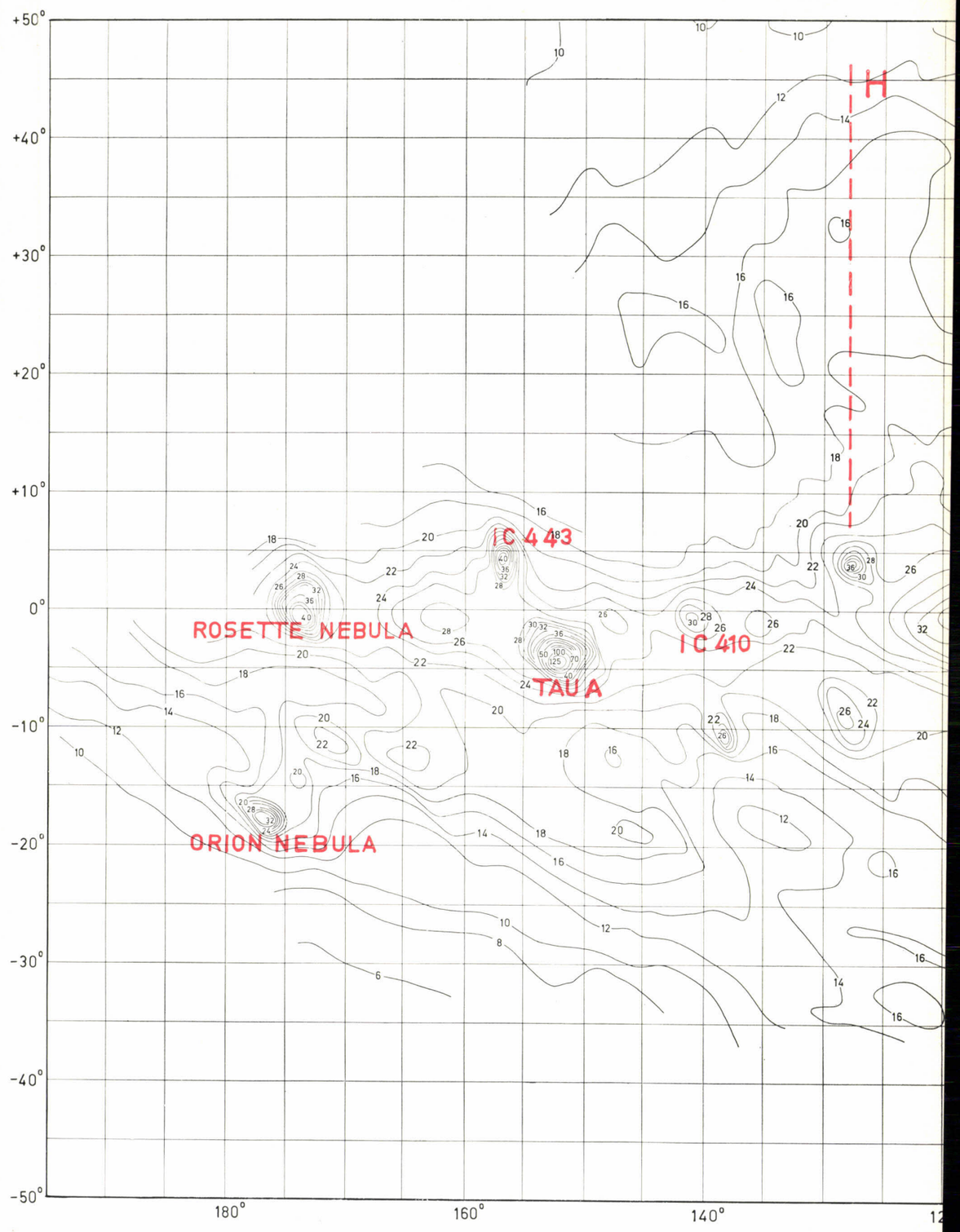


Fig. 17. The distribution of the radio brightness of part of the sky, observed at 400 Mc/s with the 25 m radio telescope at Dwingeloo. The contours represent isophots (lines of constant received power). The numbers on the contours indicate the power, the unit *) being 9.1×10^{-26} W/m² per c/s, or 1.3 °K.

The region of the sky reproduced here comprises part of the Northern Milky Way. For guidance the transparent overlay sheet is marked with some of the most familiar objects in this region of the sky; the Milky Way itself is shown in grey shading, and its plane of symmetry runs through the approximately horizontal dot-dash line. The degrees along the horizontal axis indicate the galactic longitude (l^{I}), and on the vertical axis the galactic latitude (b^{I}). In the direction of the circle *C* lies the centre of the Milky Way, as seen from our point of observation, and in the direction of *A* lies the anti-centre. The stars in our vicinity travel, in their general rotation, around the centre of the Milky Way in the direction of *B* ($l^{\text{I}} = 57°$). A large part of the Milky Way as seen from Dwingeloo never comes above the horizon. This limitation of the area of the sky accessible to observation is indicated by the line on the right. The "practical" horizon shown is somewhat higher than the theoretical horizon, owing to the fact that the observations very close to the horizon are rather inaccurate. The dashed line $\delta = 5°$ (Meas. No. 419) denotes the path described by the antenna for obtaining the recording shown in fig. 16.

Some particulars to be deduced from the radio chart are outlined below.

1) The 400 Mc/s radiation is strongly concentrated around the (symmetry) plane of the Milky Way. Further, the radiation in this plane increases very considerably towards the direction of the centre.

2) The strongest point sources, or regions with high local intensity, are situated near the plane of the Milky Way. It is known that most sources seen on the chart are large regions of hot ionized gas surrounding the bright stars. These are the so-called H II regions (ionized hydrogen), familiar in classical astronomy. Apart from the exceptions discussed below, all identified radio sources (the names are printed in red) are of this type. It has not been possible to determine the true size of any of these sources; they cover a much smaller angle than the antenna beam used in the mapping of the chart. Most of the point sources clearly distinguishable as such are relatively near the sun (at distances of less than 4000 light years). For the most part they are comparatively weak radiators, but their proximity makes them seem relatively powerful. Some of these H II sources have been found to be multiple or composite sources, either by radio investigations at even greater resolving power (G. Westerhout, B.A.N. No. 488 of 31st Dec. 1958) or by optical observations. Cygnus X, for example, consists of a dozen or more separate, although closely adjacent, H II regions.

3) Four of the radio sources, whose identity is given below, differ entirely in their nature from the kind just discussed. Unlike H II sources, these four are brighter at long waves than at short.

Cassiopeia A. Except at short waves, where the sun predominates, this source is the most powerful heavenly object in the whole radio spectrum. Following F. G. Smith's careful localization by means of radio observation, Baade and Minkowski have examined the source with the 200″ Palomar telescope. They found it to be a wispy, roughly spherical shell, consisting of somewhat cirrus-like gaseous clouds, and rapidly expanding at radial velocities of several thousand kilometres per second. The source lies about 10^4 light years away in one of the outer spiral arms of the Milky Way. It is believed to be the remnants of an exploded star, or nova. Incidentally, the isophot 300 of this source gives a very accurate picture of the transverse section at half the peak intensity of the antenna beam, since it is known that this radio source has a diameter of only $^1/_{12}$th degree and that it is very much more powerful than any other source in the immediate vicinity.

Cygnus A. Although this radio object in the sky appears to be the second in brightness, in reality it is the most powerful radio source known. This object, too, has been identified with certainty as a result of the work of Smith, Baade and Minkowski: it consists of two galactic systems in collision at a distance of 10^9 light years. If they were somewhat further away, they would not be observed with the present-day optical telescopes, but they would still easily be "seen" by the radio telescope.

Taurus A. The celebrated Crab Nebula, several times referred to in this article (fig. 2) consists of the remains of a vast stellar explosion (supernova), about 4000 light years away. As mentioned, this Nebula emits synchrotron radiation.

IC 443. This is another cirrus-like gaseous object. Its identification with the radio source seems well established, but the origin of the radiation is not yet clear.

4) Within about 60 degrees of longitude from the centre of the Milky Way the level of the total radio-wave radiation increases towards the centre to a fairly flat maximum, as mentioned under (1). It is known from observations at high and low frequencies that at least two different mechanisms contribute to this radiation. An appreciable part of the 400 Mc/s radiation observed here can be attributed to H II regions which, in the direction of view, are distributed over our entire galactic system, but the major part belongs to a background which must have an entirely different origin. The spectrum of this second kind of radiation shows an intensity that increases with increasing wavelength, rather like the radiation from the sources mentioned under (3). The cause of this non-thermal radiation has not yet been explained, although there are various reasons for assuming that it bears a relation to a general magnetic field in the Milky Way and to particles of very high velocity such as occur in cosmic radiation. This is the radiation component discovered in about 1930 by Jansky, the pioneer of radio astronomy, and which in later investigations was found to be characteristic all of galactic systems of the same type as our own. In the middle of the region from which this non-thermal radiation comes, there is another powerful H II source, Sagittarius A. This source is very close to the centre of rotation of the Milky Way.

5) Of the large galactic systems outside our own, the nearest is M 31, the Andromeda Nebula. Its position is marked on the overlay sheet. Although this object can be clearly distinguished in our recordings, it is barely visible in the chart owing to the comparatively large intervals between the isophots drawn, the more so since the same region of the sky contains a small arm of radiating material which appears to belong to our own Galaxy.

6) A remarkable feature, shown on all radio charts made at these wavelengths since the early work of Reber, is the protuberance indicated by the dashed vertical line *D*. Our chart shows more structural details of this than any of the earlier charts, owing to the higher resolving power and sensitivity of our equipment, and perhaps also as a result of the brightness distribution being different at other wavelengths.

This radio protuberance has no counterpart in the optical picture of the Milky Way, and its origin is quite unknown. It is not even certain whether it consists of a local bulge or whether it is a peculiarity in the shape of the entire Milky Way. There is another phenomenon to be seen, and this was first noticed when making the present chart: this protuberance seems to be echoed in a systematic manner by two similar but weaker protuberances in the directions indicated by the dashed lines *F* and *H*.

On the whole it can be said that our galactic system possesses a high degree of symmetry, both optically and in terms of radio waves: it has rotational symmetry and moreover possesses a mirror plane of symmetry. The protuberances referred to are a notable exception to this rule: radio waves, of relatively low intensity, are found only on one side of the plane of symmetry, up to a large galactic latitude. Further, the northern protuberances *D* and *F* appear to be accompanied by two minima, *E* and *G*, which occur at the same galactic longitude, at southern galactic latitudes. This markedly antisymmetrical distribution of the radiation is a characteristic of the radio Milky Way which, until the completion of this chart, had gone more or less unnoticed. Comparable charts made at much longer wavelengths show the effect to nothing like the same extent. Whether it also occurs at much shorter wavelengths is not known, radio telescopes not yet being sensitive enough for such observations.

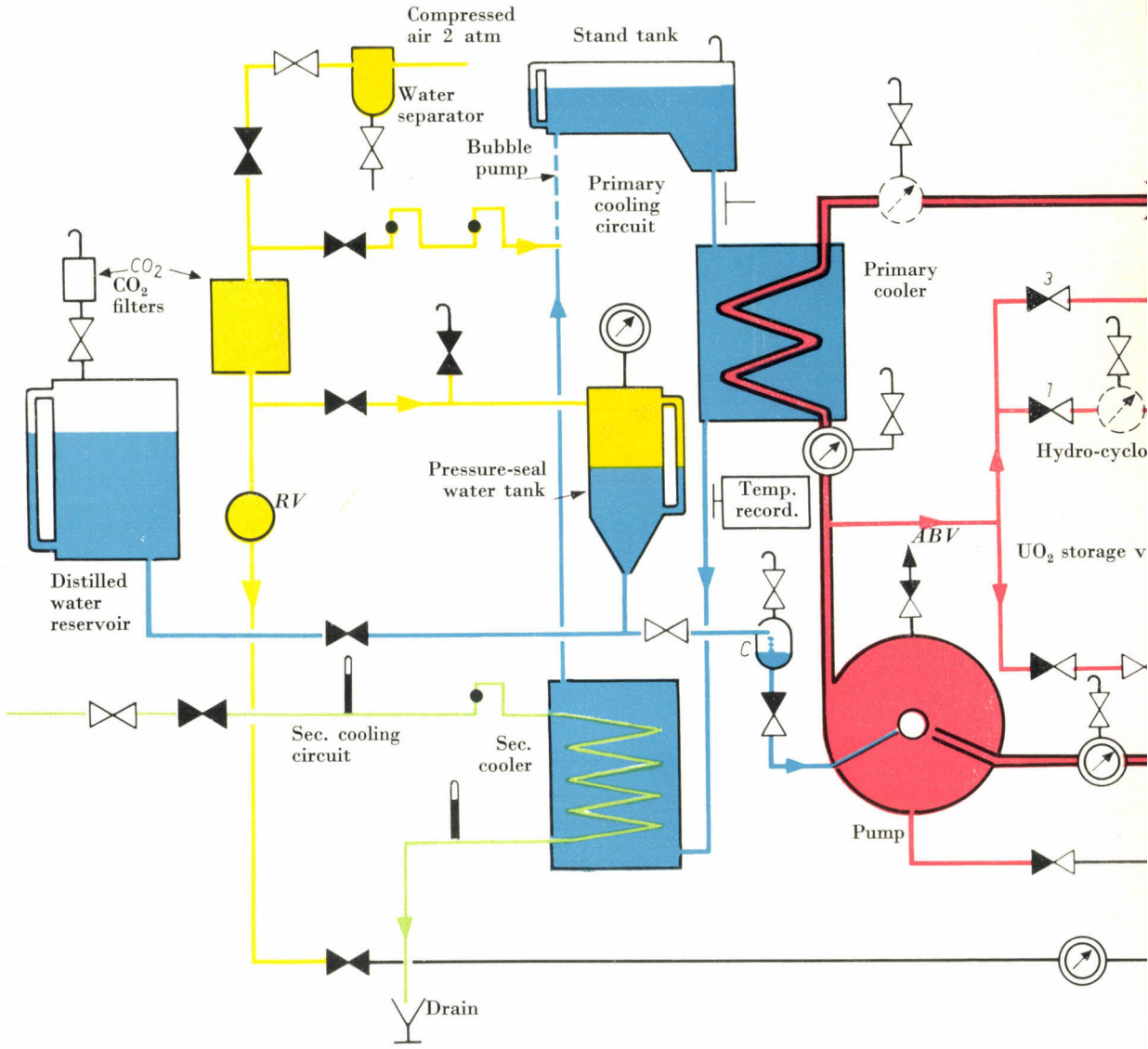
*) This value (the proportionality coefficient in the relation, found by calibration, between the output voltage and input power of the receiver system(has an estimated error of ± 15%. In this connection see the thirdarticle mentioned in [15]). The estimated error in the scale zero-point is 10 to 15 °K.

system and the associated apparent anti-symmetry of the Milky Way in relation to its "plane of symmetry", demonstrated by this chart for the first time, deserve particular attention.

A constantly recurring problem in radio-astronomical observations is the occurrence of radio interference, due to transmitters and their harmonics (e.g. in our case radio-sondes between Wales and Copenhagen, and aircraft transmitters between the West of England and South Germany) and to secondary causes, such as the ignition systems of vehicles at distances up to 5 or 10 km away. Even in Dwingeloo, which is fairly remote from busy areas, 25% of the observation time at night has been lost through interference. (During the day observations of weak sources in and beyond the galactic system are not possible because of the intense radiation from the sun.) The same difficulties are experienced elsewhere, and the subject has fortunately attracted international attention. On a proposal of the Netherlands the C.C.I.R. (Comité Consultatif International des Radiocommunications), the technical advisory body to the Post Offices of various countries, has accepted a recommendation to reserve certain frequencies for the purposes of radio astronomy. At the last meeting, concluded 21st December 1959, of the I.T.U. (International Telecommunication Union) which is the body responsible for frequency allocations, certain frequency bands were in fact allocated exclusively for radio astronomy, and other frequency bands were indicated for radio astronomy together with non-interfering services, e.g. beam-approach installations on airfields. Time will show whether this allocation will meet with the necessary international cooperation. In view of the common interest in the science of radio astronomy, which has flourished so enormously in the last ten years, this cooperation may well be expected.

Summary. The article describes a simple, stable and sensitive receiver for the frequency range from 200 to 500 Mc/s (150 to 60 cm), possessing a very low noise factor ($F = 2.5$ at 400 Mc/s), which makes possible accurate observations of the very weak radio waves received from extra-terrestrial sources. The receiver operates on the superheterodyne principle, with an intermediate frequency of 50 Mc/s. The high-frequency amplifier contains two EC 57 disc-seal triodes in cascode, the effective cathode resistance of the second valve being made very high by a simple neutralization circuit. This arrangement provides a very high gain and virtually eliminates the noise contribution from the second valve. Various measures, including the use of a constant-impedance detector and frequency stabilization of the local oscillator with a quartz crystal specially designed for observations at 400 Mc/s, make the receiver so stable that its sensitivity shows no noticeable drift for weeks on end. For calibrating the detector the I.F. amplifier contains a calibrated attenuator, adjustable in steps of 1 dB. The receiver can be given a bandwidth of 10 Mc/s or 1 Mc/s, and the time constant can be varied between 0.1 and 10 sec.

Various radio-astronomical observations are described that have been made with this receiver, particularly in conjunction with the 25 m parabolic reflector of the Radio Observatory at Dwingeloo. When this came into operation in 1956, it was the largest rotatable radio telescope in the world. The isophot chart of the sky drawn from the observations made with this radio telescope, and reproduced in this article, provides the most detailed information at present available on the structure of our Galaxy, measured at 400 Mc/s. Particularly remarkable phenomena are various tails of the general radiation from our Galaxy, and associated with these the antisymmetry of the Galaxy in relation to its "plane of symmetry." The absolute level of the measured radiation intensities is determined by calibration with the aid of a horn antenna: the radio star Cassiopeia A, the most powerful object in the radio sky at these wavelengths, gives a flux density of 56×10^{-24} W/m² per c/s.



Legend:

Pipes outlined in black belong to the reactor circuit; the others are ancillary lines for control, cooling, etc.

Red: UO_2-water suspension Green: main water
Blue: distilled water Yellow: compressed air

Valve, not on control panel

Valve operated by remote control

Valve, on control panel

Resistance thermometer

Mercury thermometer

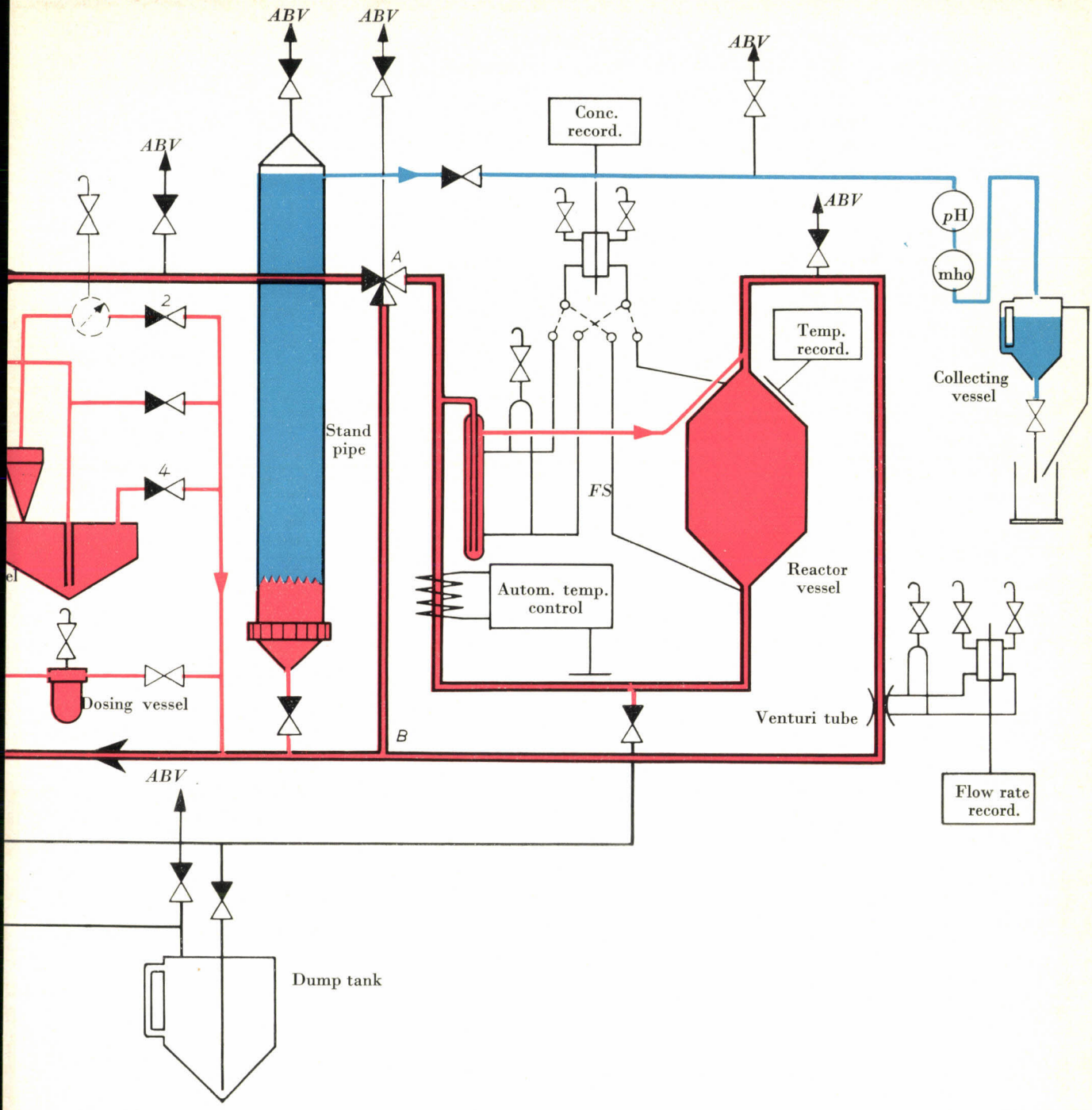
Manometer, *in situ* on reactor and also on control panel

Manometer, only on control panel

Manometer, *in situ* on reactor only

Float-type flowmeter

ABV



ectric differential pressure pick-up

r bleeder

r bleeder via bleeder vessel, into
nich no CO_2 can enter from outside

vel gauge

Fig. 21. Flow sheet of the subcritical suspension reactor installed in one of the laboratories of N.V. KEMA at Arnhem. Through the reactor circuit (see legend) a suspension of UO_2 particles in distilled CO_2-free water is kept in continuous circulation. The pump speed is not variable; the flow rate in the reactor vessel is varied by means of a 3-way valve *A* which controls the fraction of the circulated fluid flowing through the shunt pipe *AB*. The flow through the reactor vessel is measured by a venturi tube (with an electric differential pressure pick-up) and continuously recorded. Also recorded are the temperature in the reactor vessel, the temperature of the cooling water upon leaving the primary cooler and the average concentration of the suspension in the vessel. The very sensitive and stable differential pressure pick-up used for the latter purpose can be connected either to the flow simulator *FS* or to the reactor itself, as required. The automatic temperature control is upstream of the reactor vessel. By opening valves *1* and *2* and setting the hydro-cyclone in operation, a very large proportion of the UO_2 particles can be removed from the fluid and collected in the UO_2 storage vessel. The latter can be emptied by closing valves *1* and *2* and opening *3* and *4*. By means of the dosing vessel (like the UO_2 storage vessel, detachable) small known amounts of UO_2 can be added to the circuit. The flow of pressure-seal water is measured with a droplet flowmeter *C*. The acidity (*p*H) and the electrical conductivity (mho) of the water are measured as it leaves the stand pipe. If necessary, the reactor circuit can be completely drained, the UO_2 particles being first transferred to the UO_2 storage vessel and the water then drained into the dump tank. This tank can be emptied by compressed air taken via a reducing valve *RV* from the air line which also serves the bubble pump in the primary coolant circuit and the pressure-seal water line.

URSA MINOR

POLARIS

LYRA

CYGNUS

B

SCORPIUS

C

$\delta = 5°$
Meas. Nr. 419

CASSIOPEIA

PERSEUS

AQUILA

NORTH AMERICA NEBULA

SAGITTARIUS

PRACT. HORIZON DWINGELOO

ANDROMEDA

PEGASUS

+50°

+40°

+20°

0°

−20°

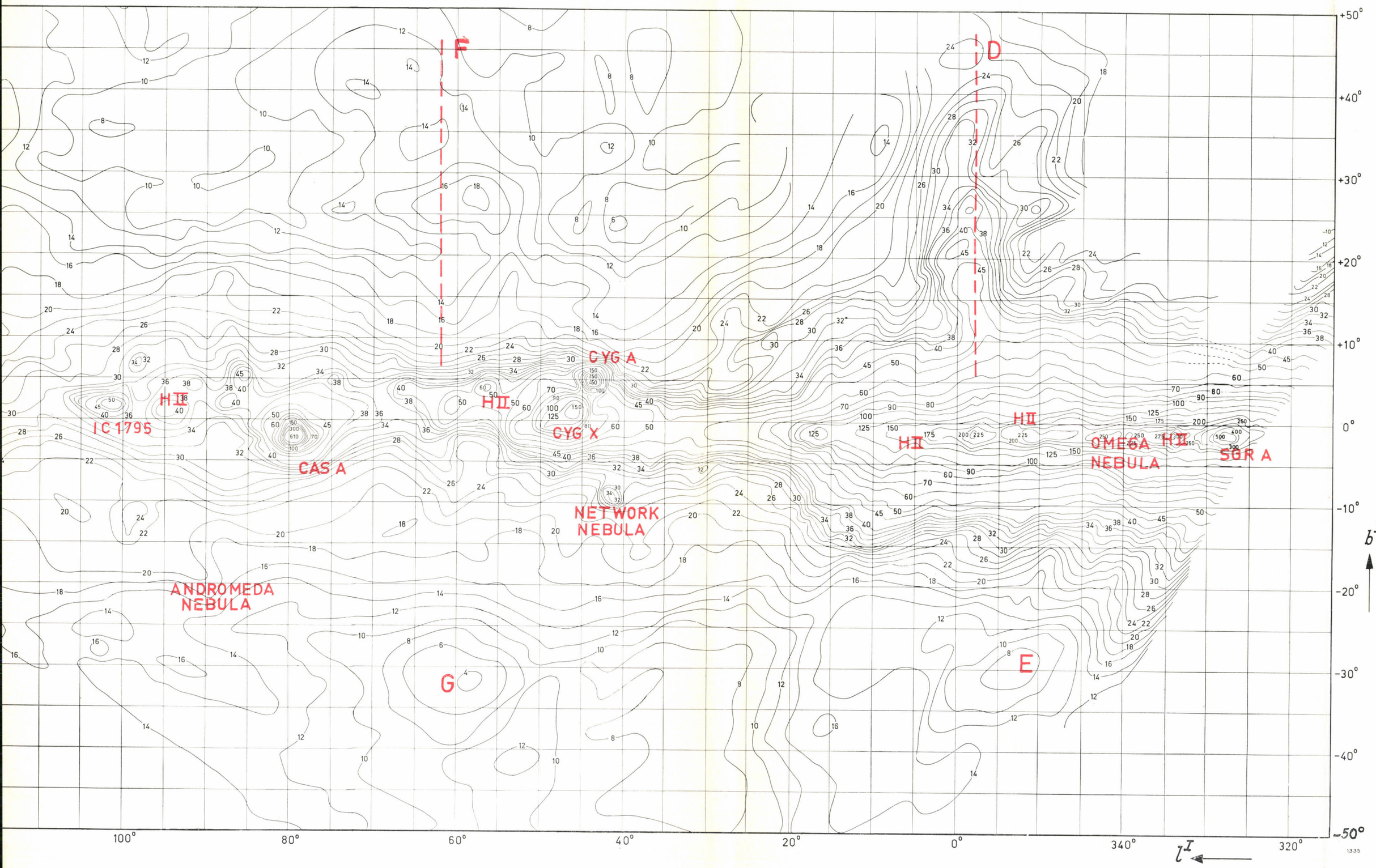
−40°

−50°

b^I

100° 80° 60° 40° 20° 0° 340° 320° l^I

1335


F
D
CYG A
CYG X
HII
IC 1795
CAS A
NETWORK NEBULA
ANDROMEDA NEBULA
OMEGA NEBULA
SGR A
G
E
+50°
+40°
+30°
+20°
+10°
0°
-10°
-20°
-30°
-40°
-50°
b^I
100°
80°
60°
40°
20°
0°
340°
320°
l^I
1335

THE FRUITS AND FOUNDATIONS OF SOLID-STATE RESEARCH

by D. POLDER.

539.2:54-16

The article below broadly reproduces the text of the address delivered by the author on 2nd Dec. 1959 upon his inauguration as extra-mural professor at Delft. With a few examples the author illustrates the significance of the solid state in electrotechnical applications. He then surveys the evolution of the theory that forms the basis of any discussion of solid-state phenomena. With Professor Polder's kind cooperation, the text is supplemented here by a concise bibliography.

"Germanium, atomic number 32, atomic weight 72.6, melting point 958 °C, specific gravity 5.35, had already been predicted by Mendeléeff as eka-silicon before it was discovered by Winkler in 1886." So reads the first sentence of the article on germanium in the sixth edition of the Dutch Winkler-Prins Encyclopaedia, published in 1950. The last sentence of the article is also worth quoting; it reads: "The element has no practical value whatsoever." Taking into account that the publication of this edition was delayed by the Second World War, it is fair to assume that this sentence broadly represents the state of affairs some fifteen years ago. Now, in 1959, more than a hundred-million transistors a year are produced from the element germanium in factories all over the world. Germanium, which for years had received only passing mention in text books of chemistry, and whose physical properties had been studied only superficially, may now justly claim to be the most thoroughly investigated, chemically, physically and technologically, of all the solids.

Germanium is a semiconductor. Before 1940 the effects occuring in semiconductors were still in many respects perplexing. The electrical conductivity of these substances was much smaller than that of metals, and yet too large to allow them to be classed as insulators. In some cases the conductivity was found to be virtually independent of temperature, and in others it increased very markedly with temperature, an effect unknown in metals [1]). Deviations from Ohm's Law were often found even at very low current densities. The worst was that the behaviour of these already somewhat strange chemical substances proved to be so sensitive to impurities. Many a right-minded physicist was therefore inclined to regard the phenomena in semiconductors with deep suspicion.

The unusual conduction properties of semiconductors had meanwhile attracted the attention of those who were engaged in the search for curious properties of matter that might be capable of practical application. The negative temperature coefficient of the resistance of oxidic semiconductors was turned to good use for limiting the heater current when switching-on a thermionic valve [2]). The non-linear electrical characteristic of certain semiconductors found application in the crystal diode, which became an invaluable aid in the detection of radar signals [3]). It was in this connection, during the war, that germanium was first studied systematically as a semiconductor. And it was the demand for practical applications that attracted to fundamental semiconductor physics the attention which many physicists had not believed it deserved.

The great moment in the history of semiconductors came in 1948 when, as the result of their work in the Bell Laboratories, Shockley, Bardeen and Brattain announced their discovery of transistor operation [4]). This discovery led to the possibility of making a piece of germanium fulfil the function of a thermionic valve. As an amplifier of electric signals the germanium transistor now rivals the valve. In itself, transistor operation is a most remarkable physical phenomenon. Moreover, it is a very well understood phenomenon, which is in no way bound to the specific chemical properties of the element germanium. In principle it can also occur in other elementary semiconductors and in semiconducting compounds.

The advent of the transistor has profoundly influenced the recent development of solid-state physics. The intensive study of all possible kinds of semiconductors in laboratories far and wide [5]) is one direct result. Seen more broadly, the existence of the transistor and associated semiconductor devices has focused attention very sharply on the many and varied technical potentialities of solid-state properties, and in particular of those properties bound up with the behaviour of electrons in solids. It would be wrong to assume, however, that the technical possibilities of solids in electronic engineering had

scarcely been recognized before the transistor came on to the scene. The extensive work done and still being done, in the Netherlands [6]) and elsewhere, on oxidic semiconductors and magnetic materials, testifies to the contrary. The applications of magnetic materials extend over the whole frequency range of electromagnetic waves covered by electronics. Research on magnetism, including fundamental research, is strongly influenced by the demand for technical applications. On this the transistor has had no bearing at all.

In a wider context than that of solid-state research, the transistor is held up as striking evidence of the truth of the assertion that those in search of phenomena that can be turned to practical use should devote themselves to pure scientific research. The first sentence of the introduction to Shockley's celebrated book on transistor physics [7]) reads: "If there be any lingering doubts as to the wisdom of doing deeply fundamental research in an industrial laboratory, this book should dissipate them." It is not difficult to enumerate aspects of the invention of the transistor that support this assertion. One can say that the transistor effect found in the element germanium was a particularly surprising one, and it was recognized for what it was only because it was discovered in the course of a systematic fundamental investigation. One can further maintain that the exploitation of the effect to a technically useful application was possible only because of the complete and fundamental understanding of the physical effect. It is more difficult to indicate *how* fundamental the research should be in an industrial laboratory [8]). There is another sentence in the introduction to Shockley's book that should not be overlooked in this connection; it reads: "But there should be no illusions about the necessity of a large measure of good luck".

Whatever one might think about the relation of science to industry, there is no denying that solid-state research has increased enormously since the discovery of the transistor, and that much pure scientific work is now being done in the large industrial laboratories. The interest in the solid state, shown by large and small industrial undertakings and by some governmental departments, appears for example from the many advertisements in American journals and magazines that hold out solid-state research to the physicist as a chance of beginning a rewarding career in industry. The advertisements are interesting, too, in that they give a fairly reliable indication of the subjects at present being studied: infra-red detectors, solid-state microwave amplifiers, thermoelectric cooling, data storage elements, high-frequency diodes, the theory of "devices", photoconductors, semiconductor technology, thin magnetic films. These are only a few of the subjects that are engaging the special attention of the electronics industry.

In describing the fortunes of a single chemical element, germanium, I have tried to give some idea of the vigorous developments that are taking place in the whole of the solid-state field. The example chosen has allowed me to emphasise the great stimulus which research has received from the demand for technical applications. Fortunately, there are many fields of solid-state physics that have grown without this stimulus. I say fortunately, not because I deplore the stimulus or regard it as secondary, but because I feel it is a good thing for scientific curiosity that it does not have to depend on such a stimulus. To give an instance I shall discuss briefly a subject which is studied mainly at the Universities.

After the last world war, research was started at the Clarendon Laboratory, Oxford University, on the absorption of microwaves in paramagnetic salts [9]), use being made of the facilities created by the wartime development of radar engineering. From the theoretical work of Van Vleck, from susceptibility measurements and measurements of specific heat at low temperatures, it had long been known that the energy levels of paramagnetic ions in crystals may lie very close together [10]). In order to determine these levels by direct spectroscopic means, electromagnetic waves of wavelength about 1 cm are needed. Microwave absorption, or paramagnetic resonance, was therefore at that time primarily microwave spectroscopy. The microwave investigations undertaken at Oxford and many other places have now taken on a much wider significance. A subject of investigation of great interest at the present time is the mechanism of relaxation; this is concerned with the rate and the nature of the processes that tend to restore thermal equilibrium between the various energy levels. Another subject at present receiving widespread attention is the use of the paramagnetic resonance technique, and of the associated method of nuclear magnetic resonance, for studying the structure of matter [11]). In research on the structure of impurities and other lattice imperfections, these resonance techniques have proved to be particularly valuable. Here we have an example of an academic application of academic research.

It is interesting to note that resonance investigations have also led to a number of technical

applications. Nuclear magnetic resonance, for instance, is used in the chemical analysis of mixtures of organic compounds, e.g. in the oil-refining industry. Another application, which I shall deal with at greater length, is based on a discovery made by Bloembergen of Harvard University. From Townes' work on the microwave absorption of gas molecules it was known that, under certain conditions, molecules in a microwave field will emit instead of absorbing radiation [12]). This is the phenomenon known as stimulated emission. Since the stimulated emission is *coherent* with the incident electromagnetic wave, the effect can be used for amplifying the incident signal. In the *maser*, which is an acronym for *m*icrowave *a*mplification by stimulated *e*mission of *r*adiation, the fundamental phenomenon of stimulated emission was specifically put to use for the first time. A necessary condition for obtaining a *nett* emission is that, of the two molecular energy levels involved, an excess of molecules should on average occupy the upper level. In thermal equilibrium the opposite is the case. The reversed occupation density required must therefore be brought about by some artifice. Bloembergen set out to see whether the maser principle could also be realized in a paramagnetic salt [13]). From the work on microwave spectroscopy a great deal was known about energy levels having energy differences of the appropriate order of magnitude. The study of paramagnetic relaxation at the temperature of liquid helium had shown that it must be possible here, too, to achieve the necessary reversed occupation, and by a relatively simple microwave technique. Once Bloembergen's ideas on the solid-state maser had become known, this new microwave amplifier was in fact successfully demonstrated at various places. The special interest shown in the maser in technical circles, in spite of the objections to working with liquid helium, is explained by the fact that masers are amplifiers possessing, in principle, very low inherent noise. For amplifying extremely weak signals from extra-terrestrial sources this is a feature of the utmost importance.

In the foregoing we have touched on two amplifiers, the transistor and the maser, both of which exploit the properties of the solid-state. This will allow us compare the principles underlying the amplification in these and other devices. The transistor, like the thermionic valve, operates in the first instance with a continuous current which is periodically varied by an alternating current signal in such a way that the signal thereby performs virtually no work. The energy delivered by the amplified signal originates in the source which supplies the continuous current. Of course, the mechanism causing the current to flow through the valve differs from that which maintains the flow of current in a transistor. In the valve the current is produced by an electric field, whereas in the transistor the current is the result of a concentration gradient of charge carriers injected in a semiconducting layer. Nevertheless, the transistor and the thermionic valve, regarded as four-terminal networks, have much in common.

The maser, on the other hand, is a true negative-resistance amplifier. To be more precise, in the maser the paramagnetic salt behaves in relation to the signal like a tuned circuit, consisting of a negative resistance, a negative inductance and a negative capacitance. In itself this combination has the same characteristics as a tuned circuit composed of positive elements, but the moment it is loaded with a positive resistance it can be used as a generator or amplifier of electromagnetic oscillations. Of course there is no question of getting something for nothing. To trace the energy source of a negative-resistance amplifier, we must look into the physical cause of the negative resistance. In the maser we find at once that the energy is supplied by the stimulated transitions of the molecule or paramagnetic ion from a higher to a lower energy level. The ingenious mechanism which sustains the reversed occupation introduces into the system the energy required for amplification.

The great difference between amplifiers of the thermionic valve and negative resistance types is that, in the first type, changes in the output circuit scarcely influence the input signal, whereas in the second type that is just what they do. Since the absence of feedback may be an important practical requirement, it is desirable to complement the negative-resistance amplifier with directional isolators or "unilines". These are network elements which transmit signals coming from one direction but not from the other. Under certain conditions, solids in thermodynamic equilibrium exhibit the non-reciprocal properties needed for constructing a directional isolator [14]). The classic example is the microwave uniline devised by Hogan which exploits the non-reciprocal permeability tensor in pre-magnetized ferrites [15]). In an analogous way one can use the non-symmetrical resistance tensor of a semiconductor located in a magnetic field [16]).

In this connection it may be noted that, in recent years, special attention has been focused on amplification by means of solid-state negative resistances. It is hoped that this will offer new possibilities for amplifying signals of exceptionally high frequencies.

A very recent example is the *tunnel diode* invented by Esaki in Japan [17]). A tunnel diode is a semiconductor built in such a way that somewhere over a layer no more than 100 Å thick the current can flow only because of the quantum-mechanical tunnel effect. Cutting-off the tunnel current by raising the applied voltage gives rise to a negative differential resistance. Another example is the parametric amplifier [18]). Like the maser, a parametric amplifier behaves in relation to the applied signal in the same way as a circuit composed of entirely negative elements. As opposed to the maser or the tunnel diode, the formal negative resistance in a parametric amplifier is not due to the physics of a resistance element. An essential feature of the parametric amplifier is that it should contain a capacitance, an inductance or some other reactive element whose value varies together with the applied voltage or current. The energy required for the amplification is supplied by an alternating-current source whose frequency differs from, and is usually higher than, the frequency of the signal to be amplified. The operation of this type of amplifier can be demonstrated by a simple calculation. It seems, however, that this calculation is not so obvious, in view of the fact that the parametric or reactance amplifier has been repeatedly discovered in the course of this century and again forgotten. In recent years it has been found that solid-state parametric amplifiers can be made which operate at microwave frequencies. The amplifier devised by Suhl, which led to the recent revival of this type of amplifier, uses the ferromagnetic resonance phenomenon [19]). An amplifier using a variable capacitance can be made with a suitable designed semiconducting diode. In this case the variable capacitor is the semiconductor itself, which is provided with a thin, poorly conducting layer whose thickness varies with the applied voltage. The present interest in parametric amplifiers springs from the fact that, like masers, they are characterized by very low inherent noise [20]).

The practical value of the properties of the solid state extends over a very wide field. In discussing a few modern applications in electronic engineering I have, of course, done less than justice to the versatility of the existing practical applications. I have chosen my examples from a restricted field, and they were all concerned, directly or indirectly, with the amplification of electromagnetic signals. The new developments in metallurgy, to mention just one instance, have not even been referred to in passing. Versatility is the characteristic of the applications and of the properties of the solid state; and versatility also distinguishes the theory of the solid state.

I should now like to touch briefly on the theory. It is not my intention to go into details of the numerous and often specialized concepts with which the solid-state theoretician is concerned. Not that I believe that solid-state theory can be discussed only in terms of elegant — or cumbersome — mathematical formulae. On the contrary, I am convinced that any bit of theory that is really well understood lends itself to an interpretation in which formulae play only a very minor part. I believe, too, that a theoretician working together with his experimental colleagues on solid-state research should always try to arrive at this sort of non-formal interpretation of the theory. I cannot discuss details of theoretical concepts here because such a discussion takes time, the time needed to bring out the nuances of a concept and to deal with the underlying assumptions and restrictions.

I should therefore prefer to say something about the *structure* and *elements* of solid-state theory [21]). Strictly speaking it is quite wrong to speak of *the* theory of solids. It might rather be asserted that there is in fact no one theory of the solid state. After all, there is good reason to believe that the fundamental laws of physics, in particular the theory of the electron and of the electromagnetic field, wave mechanics and the discipline of statistical mechanics and its associated subject, thermodynamics, are in principle capable of explaining virtually all the phenomena of the solid state. The fact that this does not turn out to be so easy in practice is due not to any limitation in the validity of the fundamental laws, but to the inadequacy of existing techniques for applying these laws to so complicated a system as a solid. Let us regard for a moment the physical phenomena of the solid state as a melting pot for the theory of solids. We throw in all the basic subjects of theoretical physics. What finally emerges is not a beautifully formed single crystal of solid-state theory but a somewhat amorphous mass studded here and there with a number of serviceable crystallites. These crystallites are the theories of the solid state. The fragments of theory show a certain interrelationship, but they by no means fit together neatly at every point. Each fragment is a model, encumbered by numerous simplifying assumptions and abstractions.

According to the nature of the models the theories of the solid state can be variously classified. One can distinguish, for instance, between macroscopic-phenomenological theories and atomistic theories.

In a macroscopic theory, the existence of electrons and atoms is ignored. A typical example is the theory of elasticity. Atomistic theories may be subclassified into lattice theories and electron theories. From a theoretical viewpoint this classification relates to the mass of the studied particles. A lattice theory is concerned with the interrelationship and motion of the *heavy* particles, the atoms, and avoids as far as possible a detailed discussion of the behaviour of electrons. A typical lattice theory is that of Born and Madelung on the cohesion of ionic crystals. It dates from 1912, the time when Von Laue and Bragg were working on the diffraction of X-rays in solids. The theory of lattice vibrations and of the specific heats of crystals was greatly advanced by the discovery, at the beginning of this century, of the elementary quantum of action h. The models of Einstein and Debye relate the quantum characteristics of the harmonic oscillator to the existence of a finite number of modes of oscillation in a solid.

The theory of the *imperfect* lattice was a relatively late development. The concept of dislocations, originally introduced to explain plastic deformation, has proved to be particularly fruitful, and its help is invoked for explaining many other phenomena in imperfect lattices. The dislocation, a partly atomistic and partly macroscopic concept postulated by Taylor in 1934, has become a most essential element of the theory of solids [22]).

The advent of wave mechanics in about 1925, and the associated theory of electron spin and Fermi-Dirac statistics, had a revolutionary effect on the theories concerning the behaviour of the *light* particles in solids, that is on the electron theories of the solid state. The modern theory of electrons in solids stands or falls with the quantum-mechanical method of description. On the one hand there is the quantum theory of ferromagnetism and paramagnetism, in which the properties of the electron spin are the focal point. On the other hand there are the wave-mechanical theories on bound and free electrons. The most representative of these is the energy band theory, evolved by Bloch and others about 1930, which is based on the wave-mechanical model of mutually-independent electrons in a periodic potential field. The band theory explains, in principle, the existence of metals, semiconductors and insulators. It leads to the concepts, so important for semiconductors, of hole and electron conduction. It also affirms that electrons in a periodic potential field have an infinitely long free path. Without this latter idea modern *transport theory* could not exist. In the theories on the transport of charge carriers in solids an essential role is played by the interaction between electrons and atom movements. This interaction is again of particular interest at the present time: the phenomenon of superconductivity, which is gradually becoming better understood, is very subtly related to it [23]).

I have tried to show that the important and new ideas that have come into theoretical physics with quantum mechanics have helped fundamentally to form the present-day theoretical picture of the solid state. It is worth noting that this picture has undergone very little change during the recent period of boom in solid-state research. The theory of solids is rather going through a period in which ideas are being refined and elaborated, and in part cast in a more quantitative mould. In this way semiconductor research has led to further detailed elaboration of the band theory and transport theory. For example, the problem of the effect of a uniaxial compression on the fine structure of the Zeeman effect of indirect excitons in semiconductors, involves the theory in a most detailed form.

It is not to be inferred from this, however, that the theory of solids has now assumed a more or less definitive shape. In the first place, we know from experience that important consequences of an existing theoretical picture may go unnoticed for many years. Transistor action is a case in point. At the time of its discovery, the theory of transistor action fitted very well into existing ideas on the solid state, but it was evolved only after experiments had drawn attention to the existence and significance of the finite recombination rate of holes and electrons in a semiconductor, on which the transistor action depends.

Further, quantitative questions still await answers. What is the magnitude of the work-function of caesium-antimonide? How large is the activation energy of a vacancy in copper? What is the band structure of bismuth-telluride? Such quantitative questions still present the theoretician with a challenge. In attempting to answer them, he cannot get very far with general considerations. Calculations on this kind of problem so often involve doubt as to the effect of the simplifying assumptions inevitably used that little confidence can be placed in the final result. It may be that electronic computers will help in this respect by establishing empirically the relative merits of the many possible methods of calculation, by comparing the results with known experimental data. Obviously, however, the computer cannot answer the question of *why* one method appears to produce better results than another.

There are also a number of very fundamental

N.V. PHILIPS' GLOEILAMPENFABRIEKEN

questions that have not yet been satisfactorily resolved. For instance, there are good experimental reasons for assuming that the mechanism of conduction in various oxidic semiconductors cannot be described in terms of the transport theory derived from the band theory [24]). Now it is true that this theory, so successful in its application to germanium and silicon, does in fact have its limitations. It assumes the existence of mutually independent electrons which are only *weakly* bound to the lattice. One may therefore ask what form the theory should take in the case of conduction with electrons which interact *strongly* with each other and with the lattice. There is an evident need here for a simple and realistic model that will throw light on this problem,

It is for the theoretician to devise such a model, a model which, as always, should be simple enough to allow of mathematical treatment, and at the same time realistic enough to demonstrate the essential nature of the phenomenon to be described. It is interesting work, in which the theoretician is time and again compelled to take into account the nuances, the assumptions and the limitations of the concepts he wishes to use. It is interesting, too, in that he may perhaps be able, very indirectly, to contribute in some measure to the creation of new applications of solid state physics.

BIBLIOGRAPHY

1) G. L. Pearson and W. H. Brattain, History of semiconductor research, Proc. Inst. Radio Engrs. **43**, 1794-1806, 1955.
2) E. J. W. Verwey, P. W. Haayman and F. C. Romeyn, Semiconductors with large negative temperature coefficient of resistance, Philips tech. Rev. **9**, 239-248, 1947/48.
3) H. C. Torrey and C. A. Whitmer, Crystal rectifiers, Radiation Laboratory Series, No. 15, McGraw-Hill, New York 1948.
4) J. Bardeen and W. H. Brattain, The transistor, a semiconductor triode, Phys. Rev. **74**, 230-231, 1948, and the book by Shockley under footnote 7).
5) See the series of articles in the transistor issue of Proc. Inst. Radio Engrs. **46**, 949-1300, 1958 (No. 6).
6) For oxidic semiconductors, see for example E. J. W. Verwey and F. A. Kröger, Philips tech. Rev. **13**, 90-95, 1951/52. For magnetic materials see: J. J. Went and E. W. Gorter, Philips tech. Rev. **13**, 181, 1951/52; J. J. Went, G. W. Rathenau, E. W. Gorter and G. W. van Oosterhout, Philips tech. Rev. **13**, 194, 1951/52; G. H. Jonker, H. P. J. Wijn and P. B. Braun, Philips tech. Rev. **18**, 145, 1956/57; J. Smit and H. P. J. Wijn, Ferrites, Philips Technical Library, 1959.
7) W. Shockley, Electrons and holes in semiconductors, with applications to transistor electronics, Van Nostrand, New York 1950.
8) See H. B. G. Casimir, Science and Industry, Philips tech. Rev. **20**, 85-88, 1958/59.
9) B. Bleaney and K. W. H. Stevens, Paramagnetic resonance, Rep. Progr. Phys. **16**, 108-159, 1953. K. D. Bowers and J. Owen, Paramagnetic resonance II, Rep. Progr. Phys. **18**, 304-373, 1955.
10) H. B. G. Casimir, Magnetism and very low temperatures, Cambridge Univ. Press, London 1940.
11) See, for example, J. S. van Wieringen, Paramagnetic resonance, Philips tech. Rev. **19**, 301-313, 1957/58; D. J. Kroon, Nuclear magnetic resonance, Philips tech. Rev. **21**, 286-299, 1959/60 (No. 10).
12) J. P. Gordon, H. J. Zeiger and C. H. Townes, The maser — new type of microwave amplifier, frequency standard and spectrometer, Phys. Rev. **99**, 1264-1274, 1955.
13) N. Bloembergen, Proposal for a new type solid-state maser, Phys. Rev. **104**, 324-327, 1956.
14) On the principle of non-reciprocity, see B. D. H. Tellegen, The gyrator, an electric network element, Philips tech. Rev. **18**, 120-124, 1956/57.
15) C. L. Hogan, The ferromagnetic Faraday effect at microwave frequencies and its applications — The microwave gyrator, Bell Syst. tech. J. **31**, 1-31, 1952.
16) See for example F. Kuhrt, Eigenschaften der Hallgeneratoren, Siemens Z. **28**, 370-376, 1954; W. Hartel, Anwendung der Hallgeneratoren, Siemens Z. **28**, 376-384, 1954; W. J. Grubbs, Hall effect devices, Bell Syst. tech. J. **38**, 853-876, 1959 (No. 3).
17) L. Esaki, New phenomenon in narrow germanium p-n junctions, Phys. Rev. **109**, 603-604, 1958.
18) See B. Bollée and G. de Vries, Experiments in the field of parametric amplification, Philips tech. Rev. **21**, 47-51, 1959/60 (No. 2).
19) H. Suhl, Proposal for a ferromagnetic amplifier in the microwave range, Phys. Rev. **106**, 384-385, 1957; H. Suhl, Theory of the ferromagnetic microwave amplifier, J. appl. Phys. **28**, 1225-1236, 1957.
20) A. van der Ziel, On the mixing properties of non-linear condensers, J. appl. Phys. **19**, 999-1006, 1948, in particular page 1006.
21) See, *inter alia*, the following textbooks: C. Kittel, Introduction to solid state physics, second edition, Wiley, New York 1957; A. J. Dekker, Solid state physics, Prentice Hall, Englewood Cliffs 1957; F. Seitz, The modern theory of solids, McGraw-Hill, New York 1940.
22) A. H. Cottrell, Dislocations and plastic flow in crystals, Oxford Univ. Press, London 1953.
23) For a synopsis of this extremely difficult subject, see: C. G. Kuper, The theory of superconductivity, Advances in Physics **8**, 1-44, 1959 (No. 29).
24) See G. H. Jonker, Analysis of the semiconducting properties of cobalt ferrite, Phys. Chem. Solids **9**, 165-175, 1959 (No. 2), and S. van Houten, Semiconduction in NiO, to appear shortly in Phys. Chem. Solids.

Summary. Principal contents of the address delivered by the author upon his inauguration as extra-mural professor at Delft. The author takes the work done on germanium to illustrate the enormous development of solid-state research. He points to the impetus received both from industrial applications (transistors, magnetic materials, oxidic semiconductors) and from scientific curiosity (maser, tunnel diode, parametric amplifier), and draws a comparison between the various amplifiers mentioned. Turning to the theory, or rather the theories of the solid state, the author considers various theories concerned with the interrelation and movement of *heavy* particles (lattice theory) and those concerned with *light* particles (electron theories, including the band theory and transport theory). Although there is as yet no comprehensive theory of the solid state, existing theories are being refined and elaborated in many fields. Reference is finally made to the fact that many important problems, both quantitative and fundamental, still await solution.

A METHOD OF GROWING DISLOCATION-FREE GERMANIUM CRYSTALS

by B. OKKERSE.

548.571:546.289

The importance of research into the properties and technology of the semiconductors germanium and silicon is evident when one considers their application in the manufacture of diodes and transistors. The article below describes the method by which single crystals of germanium entirely free of dislocations have been made in Philips Eindhoven laboratories. This article links up with a previous article in this journal, which dealt with the purification and doping of germanium and silicon.

An important aspect of solid-state research at the present time is the study of the way in which the macroscopic physical and physico-chemical properties of crystals depend on the physical imperfections or defects of the crystal lattice. Principal among these defects are: 1) the absence of atoms at sites in the crystal lattice that should be occupied (*vacancies*), 2) the presence of atoms at places that are not lattice sites (*interstitial atoms*) and 3) linear lattice defects or *dislocations*.

The presence of dislocations in the lattice of a single crystal may affect, for example, the mechanical properties, the diffusion of foreign atoms, the rate of alloying and dissolution and also, in the case of semiconductors such as silicon and germanium, the life time of injected charge carriers. Further, depending on the circumstances, the dislocation may act as a sink or as a source of both vacancies and interstitial atoms, and in this way too it may detract from the perfection of the crystal. As far as the semiconductors referred to above are concerned — other materials will be left out of account in this article — research on the relation between crystal imperfections and physical properties is important not only to fundamental science but also to technology, in connection with the manufacture of diodes and transistors.

For the purpose of such research it would be a great help to have a material in which one particular type of lattice defect is entirely absent. This requirement, however, can be fulfilled only in respect of one type of defect: since point defects are thermodynamically stable, only the absence of dislocations enters into consideration.

In this respect germanium is a promising material. In germanium crystals, grown by conventional methods, the dislocation density, i.e. the total length of the dislocation lines in one cubic centimetre, measured in cm/cm^3 or in cm^{-2}, is about 1000 times less than in the most perfect *metal* crystals, the respective orders of magnitude being 10^4 cm^{-2} and 10^7 cm^{-2}. Some time ago it in fact proved possible to make single crystals of germanium entirely free of dislocations. We shall first describe the procedure adopted and the physical principles on which it was based. We shall then discuss the methods by which the crystals made were examined for the absence of dislocations, and we shall present the arguments which lead us to believe that the crystals were indeed free from dislocations. Finally we shall comment briefly on the perfection of the dislocation-free crystals thus obtained.

Method of preparing dislocation-free germanium crystals

The method evolved for preparing dislocation-free single crystals of germanium is simply a special method of crystal-pulling. The pulling apparatus can either be used with an ordinary crucible or with the recently developed floating crucible [1]). After the rotating seed crystal has been brought into contact with the liquid germanium and the pulling process started, the diameter of the growing crystal is gradually reduced to 1 or 2 mm. This is done by gradually raising the temperature of the melt a few degrees. The new situation is thenceforth maintained over a length of growth of about 20 mm. Finally, in a similar way, the diameter is again gradually increased to the required value (*fig. 1*) and the pulling operation is continued until the crystal is the right length. The pulling rate chosen is less than about 1 mm/min, compared with 2 or 3 mm/min in normal crystal-pulling. The flow rate of the inert gas circulating around the crystal must be kept carefully constant. It is also important that there should be no particles of solid matter floating

[1]) The method of pulling a single crystal from a melt was recently described in this journal in an article by J. Goorissen, Segregation and distribution of impurities in the preparation of germanium and silicon, Philips tech. Rev. **21**, 185-195, 1959/60 (No. 7), in which references to other literature on the subject will be found.

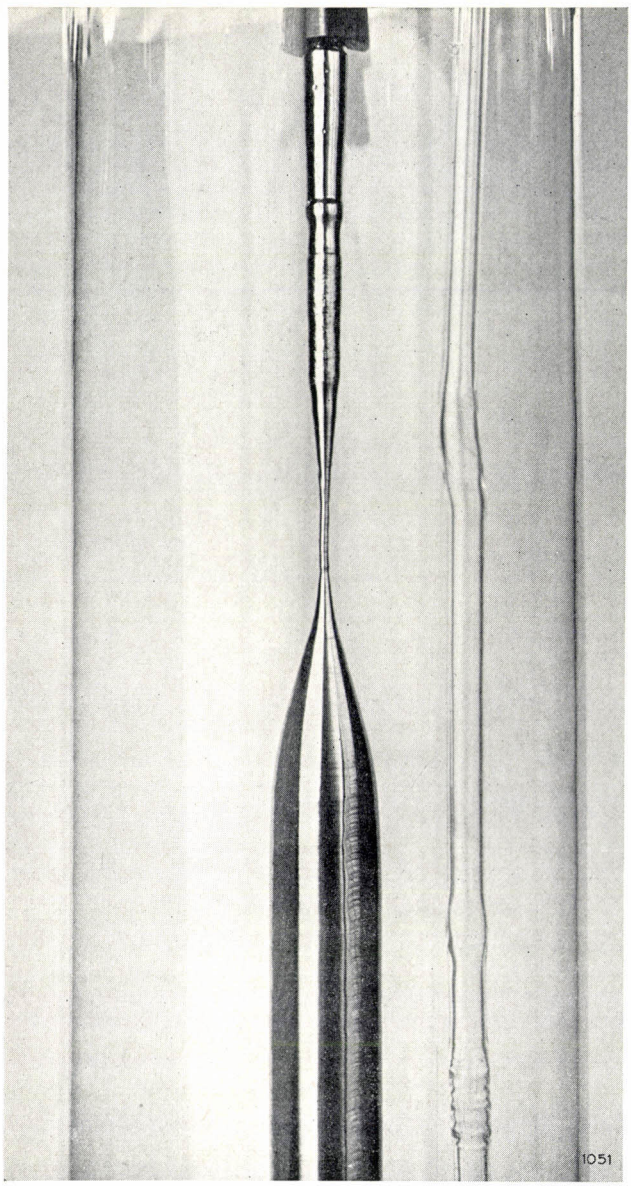


Fig. 1. Dislocation-free single crystals of germanium can be made by the pulling method if the diameter of the first part of the growing crystal is reduced to 1 or 2 mm and this diameter is maintained over a length of about 20 mm. The photograph shows from top to bottom: the seed holder, the seed, the narrowed portion, and finally the dislocation-free portion whose diameter is first gradually increased and subsequently held fixed. (The irregularities in the contours, particularly of the thin portion, are mainly due to the fact that the crystal was photographed through the wall of the glass envelope.)

on the surface of the molten germanium (e.g. germanium oxide or graphite from the crucible).

Physical considerations

The occurrence of dislocations in a crystal presupposes the presence of dislocation sources and mechanical strains. The dislocation sources may be inside the crystal as well as on the surface. A dislocation pinned at two points in the crystal can act as an internal source. A source of this kind can, in principle, give rise to large numbers of ring-shaped dislocations [2]). The nature of the surface sources is not yet clear.

With regard to the mechanical stresses required for the generation of dislocations, it may be noted that in a crystal grown by pulling from the melt these can only be caused by the presence of a temperature gradient. It can be demonstrated [3]) that a temperature gradient produces no stresses only when 1) the heat flux through a crystal is constant, 2) there are no radial heat-losses, and 3) the rate of growth is zero. Obviously, these conditions cannot be fully satisfied in practice. Fortunately, this is not necessary, since the occurrence of growth stresses leads firstly to a merely elastic deformation of the crystal lattice, plastic deformation setting in, i.e. dislocations being created, only when the stress exceeds a certain value [2]).

The reason why the method of pulling described above produces dislocation-free crystals may now be explained as follows. By initially reducing the diameter to 1 or 2 mm, and by pulling the crystal slowly, we approach the conditions that must be satisfied for the thermal stresses to be zero. Evidently, this is done in such a way that the lattice undergoes only elastic deformation and the generation of new dislocations from the existing sources is prevented. The dislocations present in the seed crystal are prevented from growing further down the crystal by making the thin section of the crystal long enough: those dislocations which intersect the solid-liquid interface at a given moment will normally grow further into the subsequently solidifying part of the crystal, but the thin neck of crystal increases the chance that they will intersect the crystal surface and thus "grow out" of the crystal (cf. *fig. 2*).

In the next part of the pulling operation, during which the diameter of the crystal is again enlarged and finally held constant at a relatively large value, the thermal stresses are certainly not small. It is difficult, however, to generate dislocations in a dislocation-free germanium crystal. Even if such a crystal is thermally "mishandled", e.g. by pulling it abruptly out of the melt during growth,dislocations are still not produced in it [4]). If, on the other hand, this experiment is done on a crystal having a dislocation density of about 10^4 cm^{-2}, the density increases to 10^6 or 10^7 cm^{-2}.

A remarkable effect is found if the melt contains

---

[2]) The mechanism of this process is described e.g. by H. G. van Bueren, Philips tech. Rev. **15,** 255, 1953.

[3]) P. Penning, Philips Res. Repts. **13,** 79, 1958. See also P. Penning, The generation of dislocations by thermal stresses, Philips tech. Rev. **19,** 357-364, 1957/58.

[4]) The same was found by Dash for silicon: W. C. Dash, J. appl. Phys. **30,** 459, 1959.

floating solid particles and these collide against the growing crystal. The part of the crystal produced after the moment of contact is found to contain a continually expanding network of dislocations, originating at the point of collision. It appears that contact with a foreign particle creates dislocation sources in the crystal surface. As the crystal goes on growing, these sources, as a result of thermal stresses, give rise to more dislocations, which later act in their turn as new sources, and so on. The existence of dislocation sources of this type has only recently become evident. From the experiment referred to in the previous paragraph, in which the

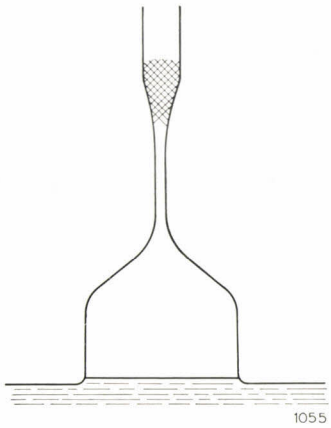


Fig. 2. In the necked portion of the crystal the dislocations as they grow with the crystal are bound to intersect the surface, where they must necessarily end; new dislocations cannot form because the thermal stresses are too small. In the lower portion, of large diameter, the thermal stresses are considerable, but now they create no dislocations because there are no longer any dislocation sources present.

crystal was suddenly cooled by withdrawing it from the melt, it appears that surface sources of this nature do not occur in a crystal grown from a melt whose surface is clean.

Evidence for the absence of dislocations

In order to demonstrate that a crystal is free of dislocations, reliable methods are needed for detecting the presence of dislocations. One of the best available methods is that based on etch pits. If a germanium crystal is cut parallel to a (111) plane and etched with a suitable agent [5]), it is found to dissolve quicker at places where a dislocation intersects the surface than elsewhere, so that after some time "etch pits" form at these places. This phenomenon is attributed to the fact that the crystal lattice around the dislocation is deformed and thus

[5]) Use can be made of the etchant described by E. Billig, Proc. Roy. Soc. **A 229**, 346, 1955, or of the etchant known as CP4, the composition of which is described in the article by Vogel et al. under footnote [6]).

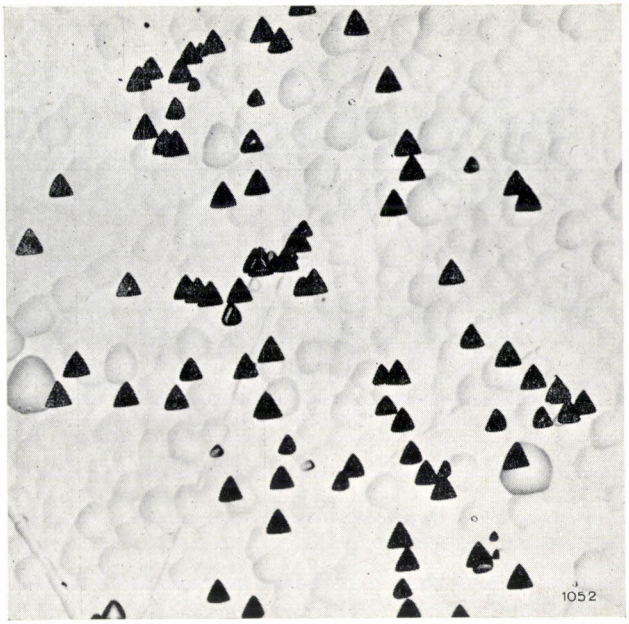


Fig. 3. If a germanium crystal is cut parallel to a (111) plane and etched with a suitable agent, the crystal dissolves quicker at the places where dislocations intersect the crystal surface then elsewhere. Each resultant etch pit corresponds to a dislocation. The sides of the triangular periphery of an etch pit are lines of intersection of the other (111) planes with the crystal surface. (Magnification 100 times.)

possesses at such locations a higher energy per unit volume. The etch pits can easily be observed and counted under a microscope (*figs. 3* and *4*).

This etching procedure is simple to use and has proved very reliable. By various methods it has been shown that in Ge, at least, all dislocations are

Fig. 4. Etched surface of three discs cut from an originally dislocation-free germanium crystal which, during the pulling process, came into contact with a piece of graphite on the melt. The disc on the right was cut at the point of collision; the two others were cut successively 5 mm further in the subsequently solidified part of the crystal. At the point of collision (see arrow) dislocation sources were evidently produced, from which stems a growing network of dislocations.

shown up in this way [6]). Various experiments carried out by us have also led to the same conclusion: the dislocation densities determined by the etching method on germanium crystals were compared with the results of two other methods. The first is based on the anomalous transmission of X-rays — an effect found only in perfect or nearly perfect crystals — and the second consists of determining the rate at which copper precipitates in a germanium crystal supersaturated with copper; this precipitation rate is governed solely by the dislocation density and the temperature (see appendix). At all the dislocation densities measurable by these methods — the lower limit of detection in our case was about 50 cm^{-2} for both methods, the upper limit for the X-ray method is some thousands of cm^{-2} — the results showed excellent agreement with those of the etch-pit method.

After etching, nearly all the germanium crystals grown by the method described above were found to be entirely free of etch pits (*fig. 5*). For this reason we believe we may conclude that these crystals are free of dislocations. In the few that were not dislocation-free, the dislocation density is several times

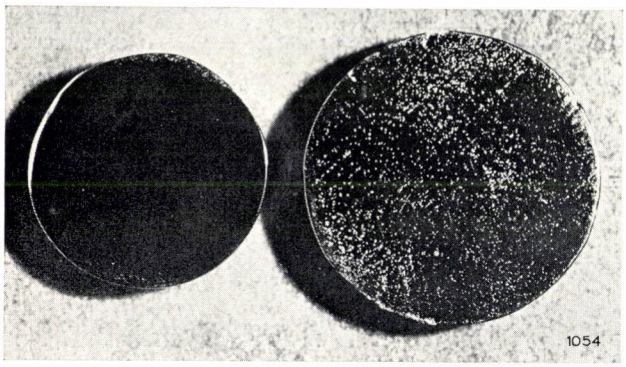


Fig. 5. Etched surface of a germanium crystal pulled by the method described (left) and of a crystal made without special precautions during the pulling process. The left crystal shows no etch pits, and this warrants the conclusion that it contains no dislocations.

10^3 cm^{-2}. The negative result in these cases is most probably due to a trivial cause, such as the presence of an overlooked piece of solid matter on the surface of the melt.

It might be asked whether the above conclusion is indeed justified, and whether it is not conceivable that the crystals appear to be dislocation-free only because the dislocations present do not happen to intersect the etched surface. The answer is that the chance of this happening in the crystals made by the method described is exceptionally small. As already noted, the thermal stresses in a crystal of normal diameter (1 to 3 cm) are considerable. As a result, the dislocation density, unless it is zero, will always be high (10^3 to 10^4 cm^{-2}) even though there was originally only one source. The chance that no single one of these dislocations will intersect the etched surface may be considered to be zero — indeed, even at a dislocation density of, say, 10 cm^{-2}, this chance would be negligible. If not a single etch pit is found in a crystal, it is therefore highly probable that the crystal is entirely free of dislocations.

In passing it may be remarked that it follows from this reasoning that, in order to demonstrate the absence of dislocations in these crystals, it is not necessary that the detection method should detect every dislocation. In practice the etch-pit method is commonly used, not so much because of its sensitivity but because it is simple.

Various investigations have shown that dislocation-free germanium crystals are not perfect. During etching, most samples do not dissolve at a uniform rate at all parts of a cross-section (see fig. 4). Moreover, the anomalous transmission of X-rays through bent crystals is found to depend strongly on position. These effects indicate that the material is not homogeneous. This is probably attributable to a non-uniform distribution of the vacancies (or vacancy agglomerates) with which these crystals are in fact supersaturated. Since there are no dislocations that can function as sinks (see appendix) the vacancies in dislocation-free crystals must diffuse all the way to the crystal surface in order to disappear. Large numbers of vacancies have insufficient opportunity to do so during the cooling of the crystal and are therefore frozen-in.

Appendix: Determining the dislocation density from the anomalous transmission of X-rays and from the precipitation rate of copper

Anomalous transmission of X-rays [7])

The transmission of X-rays incident at the Bragg angle on a perfect crystal, i.e. at the angle Θ that satisfies the Bragg law, can be described by resolving the wave phenomenon in the crystal into two components (*fig. 6*). In a direction parallel to the lattice planes that reflects the X-ray beam, both components have the character of a travelling wave; in the direction perpendicular thereto they behave like a standing wave having a period equal to the lattice spacing d. Upon emerging from the crystal the two components split into a transmitted

[6]) See F. L. Vogel, W. G. Pfann, H. E. Corey and E. E. Thomas, Phys. Rev. **90,** 489, 1953 and W. C. Dash, J. appl. Phys. **27,** 1193, 1956.

[7]) See L. P. Hunter, J. appl. Phys. **30,** 874, 1959, where references are given to earlier literature.

and a diffracted beam. The amplitude pattern of the one component (A) is now such that the minima coincide with the atomic planes and the maxima fall in between. There is therefore virtually no absorption of this component. On the other hand the maxima of the other component (B) coincide with the atomic planes, so that this component is strongly absorbed.

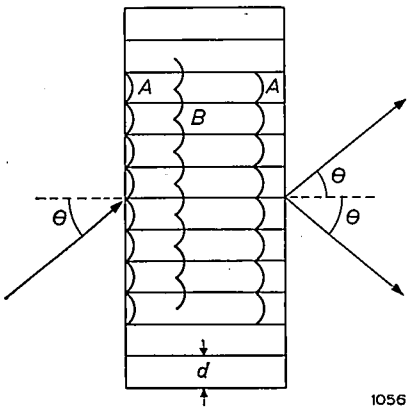


Fig. 6. Illustrating the anomalous transmission of X-rays incident at the Bragg angle on a group of parallel lattice planes of a nearly perfect crystal.

If the crystal is not perfect, e.g. owing to the presence of dislocations, then the first component, too, can suffer a degree of absorption. It is found that the intensity of the transmitted and of the diffracted beam both decrease markedly as the dislocation density of the crystal increases. The lowest dislocation density that could be demonstrated in this way with the apparatus used was about 50 cm^{-2}; it should certainly be possible, however, to lower this limit.

Calculating the dislocation density from the precipitation rate of copper

The possibility of determining the dislocation density in germanium by measuring the precipitation rate of copper is based on the interaction between dislocations and vacancies on the one hand and between the dissolved copper atoms and vacancies in the germanium on the other [8]).

Copper atoms can be found in two kinds of site in the germanium lattice; they can replace a germanium atom in the lattice (substitution) or they may occupy an interstitial site. We shall denote atoms of the first kind by Cu_s, and atoms of the second kind by Cu_i. The concentrations of these atoms are not independent of each other; there exists an equilibrium $Cu_s \rightleftarrows Cu_i + V$. In this equation V represents the vacancy produced when a substitutional copper atom abandons its lattice site for an interstitial site. The substitutional copper atoms are strongly bound to their sites (the diffusion constant is approximately 10^{-11} cm^2/s) as are the vacancies. The interstitial atoms, on the other hand, are extremely mobile (diffusion constant approximately 10^{-5} cm^2/s).

If we now saturate a germanium crystal at 750 °C with copper and then, after cooling, we raise the crystal to a temperature of 500 °C, the crystal at the latter temperature will be about 1000-fold supersaturated with substitutional copper atoms. (The solubility of substitutional copper at 750 °C and 500 °C is respectively 10^{16} and 10^{13} atoms per cm^3. The concentration of the interstitial atoms can be disregarded, since, owing to the high mobility of the interstitial atoms, it never differs significantly from the equilibrium value.) The excess substitutional copper atoms can only disappear from the crystal by first becoming interstitial. Since the vacancies thereby produced can diffuse but slowly to a site capable of functioning as a "sink" for vacancies — such as a dislocation, or the surface of the crystal — a state of quasi-equilibrium arises in which the ratio between the concentrations of the substitutional copper atoms and the vacancies is constant. The rate at which the excess copper disappears, and thus at which the solution changes from a supersaturated to the normal saturated condition, is therefore determined by the rate at which the vacancies can be removed. The latter rate depends in its turn on the number of sinks present, i.e. on the dislocation density.

Conversely, then, one can determine the dislocation density by measuring at a certain temperature T (in our case 500 °C) the time constant τ of the exponential time-variation of the substitutional Cu concentration towards its equilibrium value. In practice this is done by examining the way in which the resistivity of the crystal varies with time during the period in which the crystal is held at the temperature T. Clearly, the value of τ thus found also depends on this temperature.

It is found that the product of τ and the dislocation density N has a constant value at a given temperature. The relation between τ and $1/T$, as found experimentally by Tweet [8]), is shown in *fig. 7*. It is seen from this figure that at 500 °C the product $\tau \times N$ has roughly the value 10^9. Thus, if we find at a temperature of 500 °C a τ value of, say, 10^4, we may conclude that the dislocation density is approximately 10^5 cm^{-2}. The smallest density at which the presence of dislocations can be detected by this method is about 50 cm^{-2}, as in the previous case.

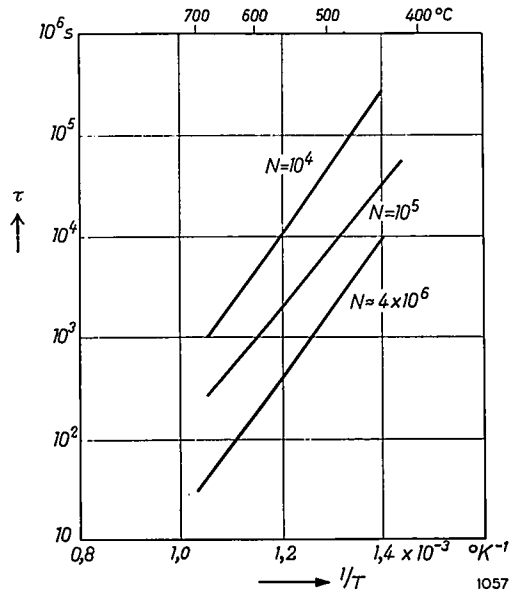


Fig. 7. The time constant τ of the rate of precipitation of copper in supersaturated germanium, plotted as a function of the reciprocal of the absolute temperature T, for three germanium crystals of differing dislocation densities N (cm^{-2}). (After Tweet [8]).)

Finally, a word about the way in which a dislocation can function as a sink for vacancies. It is known that an edge dislocation may be thought of as arising from the insertion of an extra atomic plane, the edge of which is located inside the

[8]) A. G. Tweet, Phys. Rev. **106**, 221, 1957, and P. Penning, Philips Res. Repts. **13**, 17, 1958.

crystal. Now it is very well possible that an atom at the edge of this plane may break away and occupy the site of a neighbouring vacancy site. This is particularly the case if the dislocation contains a jog: the latter then moves along the dislocation by one lattice site (*fig. 8*). A vacancy moves along a dislocation relatively fast, and the time needed to reach such a jog is equal, to a first approximation, to the time needed for a vacancy to diffuse to the dislocation.

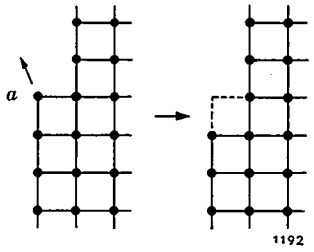


Fig. 8. An edge dislocation with a "jog" in it readily serves as a sink for vacancies. The drawing shows a part of the extra atomic half-plane of the dislocation containing a jog (left). If the atom *a* breaks away to occupy the site of a neighbouring vacancy, the jog shifts laterally along the dislocation by one lattice site as shown on the right.

Summary. Dislocation-free germanium crystals can be produced by the pulling method if the diameter of the crystal grown from the seed is initially reduced to 1 or 2 mm. This diameter is maintained over a length of about 20 mm, after which it is gradually raised to the desired value. In the thin neck of crystal thus produced the dislocations present tend to intersect the surface somewhere, where they end; thermal stresses are too small to generate fresh dislocations. The thermal stresses generally increase as the crystal diameter is increased but no dislocations are created because of the absence of sources. However, should a piece of solid material floating on the melt come into contact with the growing crystal, dislocation sources are produced in the surface. Freedom from dislocations is demonstrated by the etch-pit method. Dislocation-free single crystals still contain point defects.

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 Laboratories, Eindhoven, the Netherlands.

2706: D. J. Kroon and C. van de Stolpe: Positions of protons in aluminium hydroxides derived from proton magnetic resonance (Nature **183**, 944-945, 1959, No. 4666).

Note concerning the determination of proton positions in hydrargillite (γ-$Al(OH)_3$), boehmite (γ-AlOOH) and bayerite (α-$Al(OH)_3$). See also Philips tech. Rev. **21**, 290, 1959/60 (No. 10).

2707: F. K. Lotgering: Topotactical reactions with ferrimagnetic oxides having hexagonal crystal structures, I (J. inorg. nucl. Chem. **9**, 113-123, 1959, No. 2).

A new method is described for the preparation of polycrystalline materials with oriented crystals by reaction of the oriented grains of a strongly anisotropic ferrimagnetic with non-oriented grains of usually non-magnetic components. Using this type of reaction, for which the name "topotactical (or "topotaxial") reaction" is proposed, oriented samples of $Ba_3Co_\delta Zn_{2-\delta}Fe_{24}O_{41}$, $BaCo_\delta Zn_{2-\delta}Fe_{16}O_{27}$ and $BaCo_\delta Ti_\delta Fe_{12-2\delta}O_{19}$, which have related hexagonal crystal structures, have been prepared as well as some ferrites having the cubic spinel structures. See also Philips tech. Rev. **20**, 354, 1958/59.

2708: H. A. Klasens: The temperature dependence of the fluorescence of photoconductors (Phys. Chem. Solids **9**, 185-197, 1959, No. 3/4).

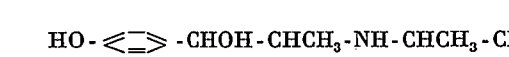
In a previous paper (No. **2650**), the effects of exciting intensity on the behaviour of a fluorescent photoconductor were discussed, using a two-state model with one level near the conduction band and occupied by an electron in the dark and one state near the conduction band and not occupied in the dark. In the present paper the effect of temperature in the stationary state for such a model is discussed. The lower level is assumed to be an activator level giving rise to a fluorescent transition, whereas the other level is attributed to a killer centre. The phosphor is excited with long wavelengths bringing electrons from the activator level to the conduction band.

It is shown that the nature of the temperature-dependence curve can be strongly affected by the capture cross-sections for electrons and holes of the two centres. A plot of $\log\{(I_0/I)-1\}$ versus T^{-1} will in general not produce straight lines as in the case of non-photoconducting phosphors, and no conclusions can be drawn from such a plot regarding the positions of the impurity levels.

The best way to determine the position of the activator level is to excite the phosphor at high intensities. A plot of $\log\{(I_0/I)^{\frac{1}{2}}-(I/I_0)^{\frac{1}{2}}\}$ versus T^{-1} will then have a slope corresponding to the thermal-energy difference between the activator level and the top of the highest occupied band.

2709: J. van Dijk and H. D. Moed: Synthesis of β-phenylethylamine derivatives, VI. Stereoisomers of 1-(4′-hydroxyphenyl)-2-(1″-methyl-2″-phenoxyethylamino) propanol-1 (Rec. Trav. chim. Pays-Bas **78**, 22-42, 1959, No. 1).

The synthesis is described of the four pairs of enantiomeric modifications of



The stereochemical structures have been correlated with those of erythro- and threo-ephedrine. Conformation analysis has been used to explain the configuration of the amino-alcohols formed.

2710: H. M. Jongerius, J. L. van Koeveringe and H. J. Oskam: Argon-xenon bands (Physica **25**, 406-408, 1959, No. 5).

Note reporting on the spectrum of a 50-50 mixture of argon-xenon. The edge of the negative glow is bright green, and contains two bands that are absent from the spectra of both argon and xenon.

2711: G. Meijer: The spectral dependence of flowering and elongation (Thesis, Utrecht, 8 June 1959).

Experiments with *Salvia occidentalis* (SDP) and *Hyoscyamus* niger (LDP) demonstrated that at least two photoperiodic reactions are involved in the process leading to a long-day effect. The main-light-period reaction is more sensitive to near infrared and blue light than to red or green light. The effect of near infrared and blue light can be antagonized by red light. The night-break reaction, promoted by red light, is nullified by a relatively short exposure to near infrared or blue light.

Experiments with various plant species on the elongation of internodes have shown that at relatively low intensities red light is more inhibitive than blue light. At relatively high intensities blue light is the most inhibiting spectral region. The inhibiting effect of red light on the elongation of hypocotyls of light-grown gherkin seedlings is antagonized by a subsequent exposure to near infrared or blue light. The inhibiting effect of red light on the hypocotyl of dark-grown gherkin seedlings is much more pronounced when the seedlings are pre-irradiated with white or blue light.

2712: J. Brug, R. J. E. Esser and G. B. Paerels: The enzymic cleavage of N-acetylneuraminic acid (Biochim. biophys. Acta **33**, 241-242, 1959, No. 1).

Under the influence of an enzyme in a filtrate of *Vibrio cholerae*, both natural and synthetic N-acetylneuraminic acid is broken down to N-acetyl-d-mannosamine and not to N-acetyl-d-glucosamine, as hitherto supposed. Under the experimental conditions no epimerization of the N-acetylamino sugars occurs. This result therefore forms a confirmation of the structure of N-acetylneuraminic acid proposed by Comb and Roseman.

2713: A. J. W. Duijvestijn: On the transition from superconducting to normal phase, accounting for latent heat and eddy currents (IBM J. Res. Devel. **3**, 132-139, 1959, No. 2).

A rigorous solution is given for the superconducting transition of a semi-infinite slab held at a point below the critical temperature T_c when a constant magnetic field above the critical value H_{oc} is applied. The solution accounts for both the absorption of latent heat and the dissipation of eddy-current heat during the transition. A numerical example is calculated for the case of constants close to those of tantalum.

2714: P. Jongenburger and C. W. Berghout: A null coil magnetometer (Appl. sci. Res. **B 7**, 366-378, 1959, No. 5).

A magnetometer has been developed in which the saturation-magnetization of small samples of ferromagnetic materials can be measured. The sample, which is situated at the centre of a small coil, is placed in an inhomogeneous magnetic field. The current through the coil is adjusted to such a value that the forces on coil and sample balance each other. Under certain conditions the magnetic moment of the sample then equals the magnetic moment of the coil; these conditions are discussed. The method is suitable for use in ordinary laboratory magnets and in solenoids, also at low temperatures. With a second version of this magnetometer it is possible to measure para- and even diamagnetism.

2715: W. F. Schalkwijk: A simplified regenerator theory (Trans. ASME **A 81**, 142-150, 1959, No. 2).

Most regenerators encountered in practice have simple linear or exponential temperature fields in their central parts. This leads to the possibility of a simplified representation of the regenerator efficiency with sufficient accuracy for all practical cases. In the important "symmetrical" case, in which the parameters have the same value in both directions of gas flow, the efficiency can be represented by a single curve. This case is treated extensively in this paper. The results for the general case are

given in an appendix. Here also a reduction is achieved of the number of parameters needed to represent the efficiency.

2716: J. F. Schouten: De beoordeling van perceptieve en selectieve belasting (T. soc. Geneesk. **37**, 393-395, 1959, No. 10). (Appraisement of perceptive and selective tasks; in Dutch.)

In industry, personnel are less and less used as instruments of mechanical energy and more and more as instruments of selection. Analysis of human selective ability is therefore desirable. It is necessary to study this ability in relation to working tempo, education and fatigue, both during the learning period and during the stabilized work-process. An example of a test to facilitate this analysis, developed in the Institute for Perception Research (Eindhoven), is described. The test is concerned with the fitting of pegs in holes, and gives some insight into the general way in which such investigations can be pursued.

2717: O. Reifenschweiler: A suitable tritium carrier for gas-discharge tubes (Proc. 2nd United Nations int. Conf. on the peaceful uses of atomic energy, Geneva 1-13 Sept. 1958, Vol. 19, pp. 360-362).

Tritium is a radioactive substance with attractive properties for use as an ionizing agent to facilitate the triggering of gas discharges. In order to introduce the tritium into the discharge tube the author has devised a method in which the tritium is absorbed in titanium powder. A suspension of the powder in a binder medium is painted with a brush on a suitable surface. To preclude that the electrons emitted are absorbed by the powder itself the particles of Ti must be extremely small. The manner of preparing such a fine powder (particle size ~ 200 Å) is described.

2718: J. L. Meijering and M. L. Verheijke: Oxidation kinetics in the case of aging oxide films (Acta metallurgica **7**, 331-338, 1959, No. 5).

The cubic relationship found by Hauffe and Kofstad for the rate of oxidation of Cu_2O in the 800-1000 °C range is confirmed, and also the abnormally low temperature coefficient of the rate constant. Special experiments show that aging effects in the CuO layer must be responsible for this behaviour. Calculations of the oxidation kinetics are given for the case when the permeability (or its reciprocal) of the oxide formed changes exponentially with age.

2719: K. W. van Gelder: Fabricagebeheersing, II. Het aanhouden van toleranties (Sigma **5**, 33-38, 1959, No. 2). (Process control, II. The meeting of tolerances; in Dutch.)

Continuation of No. **2702.** This second article deals with problems associated with the meeting of the specified tolerances, both as regards the production line itself and the inspection. In conclusion an attempt is made to set up a usable relation between quality and production costs.

2720: J. J. Arlman and H. N. L. Hoevenaar: Een niveaudetector voor de praktijk (Ingenieur **71**, Ch. 8-Ch. 9, 1959, No. 19). (A practical level detector; in Dutch).

For the determination of the liquid level in closed cylinders a simple apparatus was designed, consisting of a radioactive Cs-137 source holder and a transistorized monitor to measure the gamma radiation. This instrument provides a new method for a quick control of CO_2 fire protection installations on shipboard, which means that the time-wasting demounting, weighing and remounting of the cylinders is done away with. The simple principle permits of many applications by a mere adjustment of the source holder to the special situation considered.

2721: B. Combée and P. J. M. Botden: De mogelijkheden van teleröntgendiagnostiek, zoals bedreven in Jean Talon Hospital, Montreal (J. belge Radiol. **42**, 79-86, 1959, No. 1). (The possibilities of X-ray tele-diagnostics as practiced at the Jean Talon Hospital, Montreal; in Dutch.)

The X-ray image intensifier, together with a simply-operated television camera, makes X-ray diagnostics possible in daylight. The shielding of the operator is no longer a problem and the dose received by the patient is reduced. The installation in the Jean Talon Hospital in Montreal is described.

2722: W. Verweij and M. H. A. van de Weijer: New sodium lamps (Commission internationale de l'éclairage, 14th session, Brussels 1959).

A new design of sodium lamp is described adopting a single evacuated outer bulb around the discharge tube. The physical background to the dimensioning of these lamps is discussed. Despite its simple construction, which is considerably lighter than its predecessors with their separate vacuum jackets, a satisfactory thermal stability is ensured

by choosing the right dimensions and rare gas filling. Various features combine to give a good life performance. The use of a non-discolouring glass and depressions in the discharge tube wall which ensure a uniform distribution of the sodium and at the same time prevent the formation of sodium mirrors, improve the lumen maintenance of the new lamp as compared with the older types. Ignition voltage trouble due to rare-gas disappearance is avoided by the selection of a suitable rare gas filling. In addition to these new lamps, which are interchangeable with the older types of 45, 60, 85, and 140 W lamps using separate vacuum jackets, some experimental models for a loading of approx. 250 W are discussed.

A 12: A. Rabenau and P. Eckerlin: Zur Frage der Existenz von Verbindungen mit K_2NiF_4-Struktur (Z. anorg. allgem. Chem. **303**, 103-104, 1960, No. 1/2). (Concerning the existence of compounds with K_2NiF_4 structure; in German.)

Remarks on a paper by Wagner and Binder concerning the existence of compounds with K_2NiF_4 structure, which does not depend on the existence of a corresponding cubic perovskite.

A 13: G. Arlt: Halleffekt-Vierpole mit hohem Wirkungsgrad (Solid-State Electronics **1**, 75-84, 1960, No. 1). (High-efficiency Hall-effect fourpoles; in German.)

The upper efficiency limit of the Hall gyrator with four electrodes is 0.17. To enlarge the efficiency one has to solder more than one polarizing-current circuit and Hall circuit to the same Hall specimen. This holds for all types of Hall devices such as generator, gyrator, isolator, circulator, etc. Methods are given for the calculation of the highest possible efficiency and other interesting properties. Gyrators and isolators are proposed for audio, high and ultra-high frequency applications. Series-parallel connexion of gyrator and transformator yields a low-loss uniline.

A 14: A. Klopfer and W. Ermrich: Eigenschaften von Getter-Ionenpumpen (Vakuum-Technik **8**, 162-167, 1959, No. 6). (Properties of getter ion pumps; in German.)

The characteristic properties of getter ion pumps are reviewed, e.g. pumping capacity and pumping speed. In order to obtain low pressures it is essential to have information available on various gas-getter reactions and on desorption processes. Such mechanisms determine the ultimate pressure and the efficiency of getter ion pumps.

A 15: I. Maak and A. Rabenau: Zur Kenntnis des LiP (Angew. Chemie **72**, 268, 1960, No. 7/8). (On the compound LiP; in German.)

Note reporting three methods of preparing LiP.

A 16: P. Eckerlin, C. Langereis, I. Maak and A. Rabenau: Über $LiPN_2$ (Angew. Chemie **72**, 268, 1960, No. 7/8). (On the compound $LiPN_2$; in German.)

Note on the preparation, chemical behaviour and crystal structure of $LiPN_2$.

A 17: S. Scholz: Untersuchungen zum Einfluß des Peltier-Effektes auf Widerstands-Schweißungen (Z. angew. Physik **12**, 111-117, 1960, No. 3). (Investigation into the influence of the Peltier effect on resistance welding; in German.)

In resistance welding of two different metals the Peltier effect gives rise to welds of varying depth according to the direction of the current pulses. The magnitude of the effect is estimated: for the couple gold-palladium the Peltier heat is about 10% of the electrical energy used for the welding. Measurements confirm this result. (See also Philips tech. Rev. **20**, 188-192, 1958/59.)

A 18: A. Rabenau, A. Stegherr and P. Eckerlin: Untersuchungen im System Tellur-Thallium (Z. Metallkunde **51**, 295-299, 1960, No. 5). (Investigation of the system tellurium-thallium; in German.)

Investigations of the system tellurium-thallium were made by thermal analysis, metallographical and X-ray examination. Besides the congruently melting γ-phase and the peritectic compound TlTe a new peritectic compound with the composition Tl_2Te_3 was found. The composition range of the γ-phase could be fixed. The crystallographic data of the intermediate phases were determined.

VOLUME 21, 1959/60, No. 12 pp. 349-376 Published 24th October 1960

Philips Technical Review

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

AN ACOUSTIC SPECTRUM ANALYSER WITH ELECTRONIC SCANNING

by D. J. H. ADMIRAAL *).

534.441.2

In 1957 the Institute for Perception Research was founded at Eindhoven. Here, under the direction of Professor J. F. Schouten, members of the Eindhoven Technische Hogeschool and Philips are carrying out joint investigations in the fields of human perception and reactions and information processing.

It is intended to publish from time to time in this Review articles dealing with the work of this Institute. The first of these articles describes the way in which an acoustic spectrum analyser, originally developed in Philips Research Laboratories many years ago, has been modernized by means of semiconductor diodes and transistors.

For the frequency analysis of rapidly varying sounds, such as speech, a device was developed in Philips Research Laboratories round about 1940 which displays the Fourier spectrum on the screen of a cathode-ray tube[1]). This acoustic spectrum analyser is now in use at the Institute for Perception Research where, having been improved in various respects, it is providing valuable service in such fields of investigation as synthetic speech. For this purpose the instrument has proved just as indispensable as a cathode-ray oscilloscope in general electronic work. Some of the improvements referred to will be dealt with in this article.

First we shall briefly describe the operation of the acoustic spectrum analyser in its original form. The frequency-analysing system consists of 79 tuned circuits. Circuit 1 is tuned to 86.5 c/s, circuit 79 to 7440 c/s. The ratio between the resonant frequencies of every two successive circuits is $1 : \sqrt[12]{2} \approx 1 : 1.06$, i.e. the difference is one semitone. The tuned circuits, each with a resistor in series, constitute a series of filters ($F_1, \ldots F_{79}$, *fig. 1*). The Q of the filters being chosen at $Q = 16$, the bandwidth of each filter is also equal to a semitone. The 79 pass bands are thus contiguous and together cover the major part of the range of speech frequencies. The filter inputs are connected in parallel to the output of an amplifier A, to which the signal v_i to be analysed, originating for example from a microphone, is applied. Connected to the output of each filter is a peak-voltage rectifier; this delivers a direct voltage V_k ($k = 1, \ldots 79$) which is a measure of the sinusoidal component of the sound spectrum covered by the kth filter.

By means of a rotary switch S_1 each of the 79 direct voltages V_k drives in turn a push-pull modulator M. The signal voltage v_s is therefore successively equal to V_1, $V_2, \ldots V_{79}$, V_1, and so on. The push-pull modulator is fed with a voltage obtained from the generator G, whose frequency is about 50 kc/s. Depending on the magnitude of the signal voltage v_s, the modulator passes a smaller or larger fraction of the generator voltage, producing on the oscilloscope O a vertical deflection of corresponding amplitude. The horizontal deflection is produced by the voltage divider B and the rotary switch S_2, which turns synchronously with S_1. The switches rotate at 25 revolutions per second, so that all 79 filters are switched in at intervals of $\frac{1}{2}\frac{1}{5}$th second.

If a purely sinusoidal voltage, whose frequency f lies in the middle of one of the 79 pass bands, is applied to the input of the signal amplifier, the signal voltage v_s to the modulator will have the waveform shown in *fig. 2a* (the switch S_1 momentarily breaks circuit between each two successive contacts, caus-

*) Institute for Perception Research, Eindhoven.

1) H. G. Beljers, A recording apparatus for the analysis of the frequency of rapidly varying sounds, Philips tech. Rev. 7, 50-58, 1942.

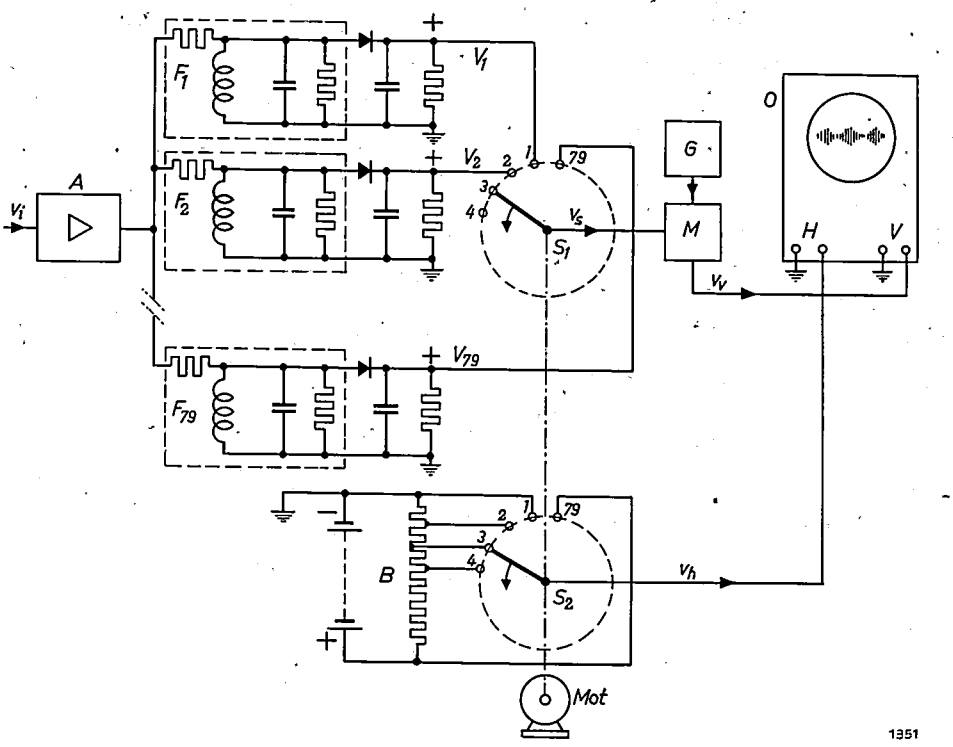


Fig. 1. Basic circuit diagram of Beljers' acoustic spectrum analyser [1]). *A* signal amplifier. $F_1, \ldots F_{79}$ filters which divide the frequency range from 86.5 to 7440 c/s into 79 consecutive frequency bands, each a semitone in width, and each provided with a peak-voltage rectifier. *G* generator (frequency approx. 50 kc/s). *M* push-pull modulator, which passes more of the generator voltage the higher is the signal voltage v_s; this signal voltage is the rectified output of the filters, which are consecutively connected to *M* via the rotary switch S_1. The output voltage v_v of *M* causes the vertical deflection on the oscilloscope *O*. The horizontal deflection is produced by a pulsed voltage v_h (see fig. 2*c*) derived from the voltage divider *B* via the rotary switch S_2. Both switches, S_1 and S_2, are mounted on the same shaft, driven by the motor *Mot*.

ing v_s to drop momentarily to zero). It can be seen that, in addition to the filter in whose passband the signal frequency f lies, the neighbouring filters also deliver a voltage. The resultant vertical deflection voltage v_v from the modulator is thus as shown in fig. 2*b*. The horizontal deflection voltage v_h is shown in fig. 2*c* (switch S_2 also breaks circuit between successive contacts; for simplicity it is assumed here that there are only 17 filters). The oscillogram (spectrogram) takes the shape shown in fig. 2*d*, which recalls the deflection pattern produced by a vibrating-reed frequency meter.

The fact that a sinusoidal signal does not produce a single-line spectrum, but a spectrum as shown in fig. 2*d*, limits the resolving power of this spectrum analyser. A discussion of this subject, and its relation to the *Q* of the filter circuits and the time constant of the smoothing circuits, will be found in the article mentioned under [1]).

The principal modification of the original apparatus is the replacement of the rotary switches by electronic switches. These have the advantage of being entirely free from mechanical noise, a point of particular importance in acoustic experiments. Furthermore, they preclude troubles due to bad contacts, which are always a drawback of rotary switches in the long run.

A photograph of the acoustic spectrum analyser in its present form is shown in *fig. 3*.

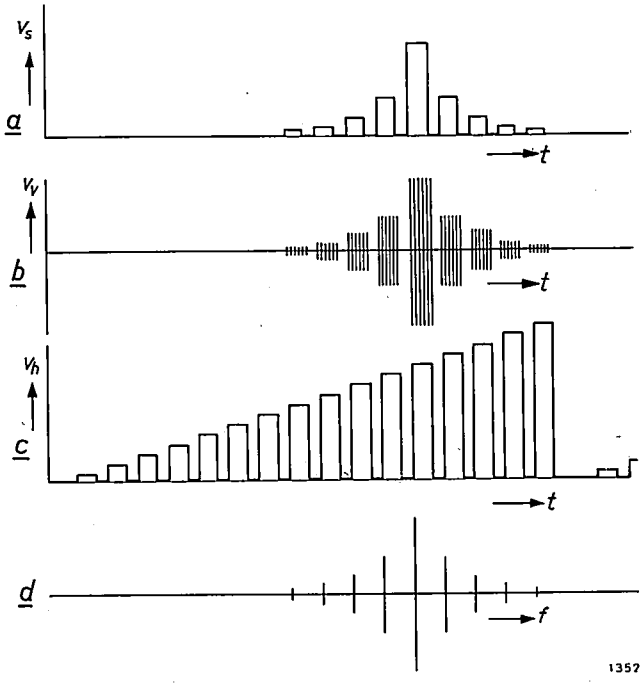


Fig. 2. *a*) Form of the signal voltage v_s in fig. 1, for a sinusoidal input signal.
b) Corresponding form of the output voltage v_v from the modulator.
c) Pulsed sweep voltage v_h (for simplicity shown here for 17 filters instead of 79).
d) Form of the oscillogram. At the peak of each pulse (*c*), v_h is constant for a moment, during which time a vertical line is traced, the length of which corresponds to v_v in (*b*).

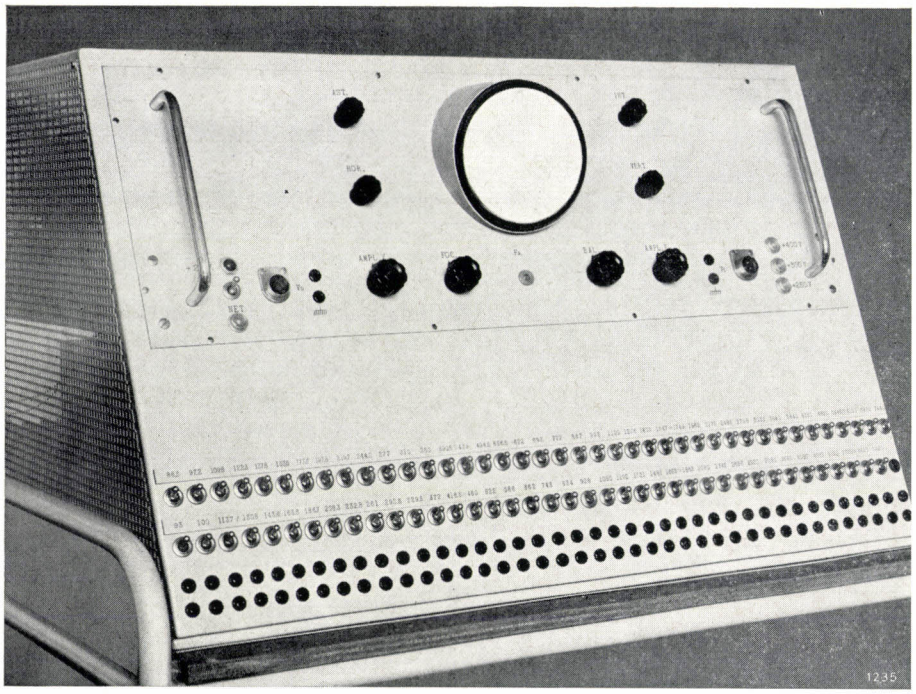


Fig. 3. The Beljers acoustic spectrum analyser in its present form at the Institute for Perception Research at Eindhoven.

Replacement of the rotary switch S_1

Each contact of the rotary switch S_1 is replaced by a diode switch, driven by a transistor circuit. The principle of the diode switch is illustrated in *fig. 4a*. In the state represented, the cathode of the diode D is at a positive potential in relation to the anode, and therefore the diode passes no current. The output voltage V_o of the circuit is then $R_o/(R_o + R)$ times the input voltage V_i. If, on the other hand, the cathode is earthed (for the present we assume this to be done by the switch Q), the diode then constitutes a virtual short-circuit across the resistor R_o, so that $V_o \approx 0$.

In principle, 79 of these electronic switches can replace the rotary switch S_1 in fig. 1. V_i is the rectified voltage V_k of one of the 79 filters, and the output V_o represents the voltage v_s that drives the modulator. (In this case, since $R = R_o$, v_s is not equal to V_k but to $\frac{1}{2}V_k$.) This means that measures must be taken to raise the cathode potential successively to a positive value in each of the 79 diode circuits.

Before discussing these measures, we must elaborate somewhat on the circuit of fig. 4*a*. In the state when the cathode is earthed the diode is not a complete short-circuit across R_o, and therefore v_s is not entirely zero. The result of this would be a spurious background in the spectrogram. The residual value of v_s can be reduced to zero, however, by instead of merely earthing the cathode of D, giving it a small negative bias — large enough to achieve the object, but not too large since that would make the residual value of v_s negative and again

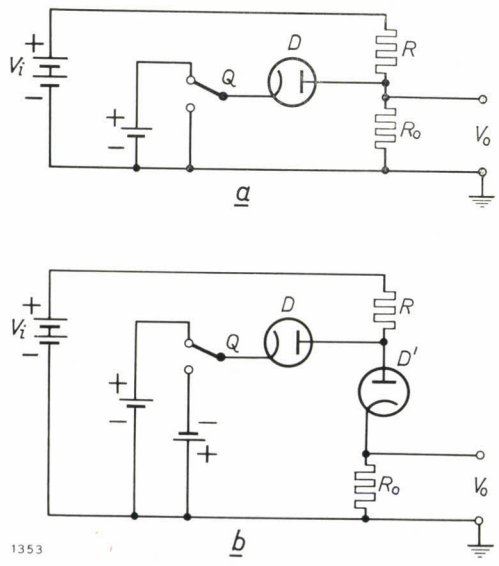


Fig. 4. *a*) When the cathode of diode D is positively biased at a potential higher than the potential drop across R_o, the diode is cut off. The output voltage V_o is then $R_o/(R_o + R)$ times the input voltage V_i. If the cathode is earthed, D constitutes a virtual short-circuit across R_o, and V_o drops almost to zero. *b*) By adding the diode D' the voltage V_o — when D is conducting — can be made exactly zero with the aid of a negative bias on the cathode of D, the value of which is not critical.

give rise to unwanted deflections. Unfortunately the appropriate bias is not identical for all diodes of the same type. Nevertheless, it is possible to make do with a single bias-voltage source for all 79 electronic switches by connecting a diode D' in series with R_o (fig. 4*b*). This prevents current flowing through R_o in the wrong direction, and as a result the residual value of v_s can no longer go negative. This allows us to raise the common negative bias to a point where the voltage v_s drops to zero in that electronic switch where it has the highest positive residual value; in the other switches the negative bias is higher than is necessary, but this can do no harm because of the presence of the diodes D'.

Fig. 5 shows a three-fold arrangement of the circuit in fig. 4*b*, provided with one common resistance R_o. The negative bias is applied to points Q_1, Q_2 and Q_3. If these points are given successively a certain positive potential, then v_s becomes successively about $\frac{1}{2}V_1$, $\frac{1}{2}V_2$ and $\frac{1}{2}V_3$.

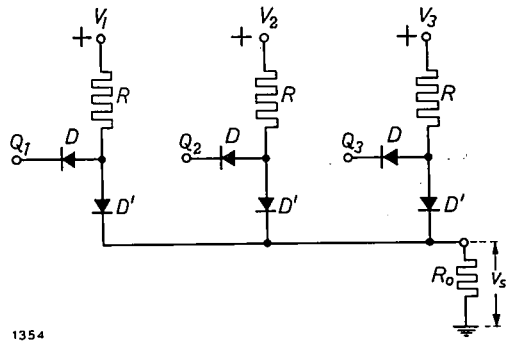


Fig. 5. The circuit of fig. 4*b*, now with three inputs and one common output resistor R_o. The diodes are semiconductor diodes.

To produce this circuit with 79 inputs we need 2×79 diodes. This makes the use of semiconductor diodes, as in fig. 5, highly desirable because of their much smaller heat development and because they dispense with the need for heater-current leads.

In order to produce the various sinusoidal components in their correct ratios in the spectrogram we must ensure that the ratio $v_s : V_k$ is independent of the applied voltage V_k. If the diodes D' were ideal (resistance R_i in the forward direction $= 0$, resistance R_{inv} in the inverse direction infinite), this condition would be fulfilled, because in that case $v_s/V_k = R_o/(R_o + R) =$ constant. In practice, however, we must replace R in this expression by $R + R_i$, and R_o by the parallel arrangement of R_o and $R_{inv}/78$. The resistance R_i depends on the current through the diode (and hence on R), and R_{inv} depends on the voltage across the diode. To minimize the effect of R_i we must make R much larger than the value of R_i, i.e. $R \geqq 100$ kΩ for a germanium diode, type OA 74, and $R \geqq 220$ kΩ for a silicon diode, type OA 202. To prevent v_s becoming much smaller than V_k, we make, as mentioned above, $R_o = R$, that is to say $v_s \approx 0.5\ V_k$.

To neutralize the influence of R_{inv} we must ensure that $\frac{1}{78}\ R_{inv}$ is much larger than R_o. With $V_k =$ 10 V the inverse voltage across 78 diodes D' is approx. 5 V. At this voltage, 78 germanium diodes of type OA 74 in parallel have a resistance of only about 50 kΩ. This is appreciably less than R_o, which must be at least 100 kΩ. In this respect silicon diodes, type OA 202, are much better: 78 of these in parallel at 5 V have a resistance of approx. 70 MΩ, which is very much larger than $R_o = 0.22$ MΩ. For this reason we decided to use silicon diodes, with $R = R_o = 220$ kΩ.

Fig. 6 shows the relation $v_s = f(V_k)$ measured for the above-mentioned arrangement. The relation is virtually linear for values of V_k larger than 0.3 V. The required proportionality between v_s and V_k is adequately approached if a bias of 0.3 V is added to V_k.

In order that the diode circuits of fig. 5 — provided with 79 inputs — shall each in turn pass the rectified filter voltages V_k to the output (with a reduction factor of approx. $\frac{1}{2}$), the points $Q_1, Q_2, \ldots Q_{79}$ must receive successively a specific positive voltage, and during the rest of the time remain at the negative potential mentioned above. This is done automatically by means of a flip-flop circuit connected to each diode D; this arrangement is illustrated in *fig. 7*. The 79 flip-flop circuits are identical and are coupled one with the other to form a ring circuit. They are supplied with a direct voltage of approx. 12 V.

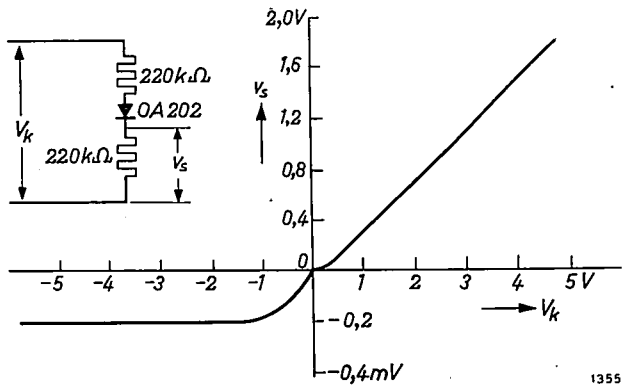


Fig. 6. Output voltage v_s as a function of the input voltage V_k for the circuit on the left, using a silicon diode OA 202. In the forward direction the relation is linear for $V_k > 0.3$ V. (Note the different scales of the positive and negative v_s axes.)

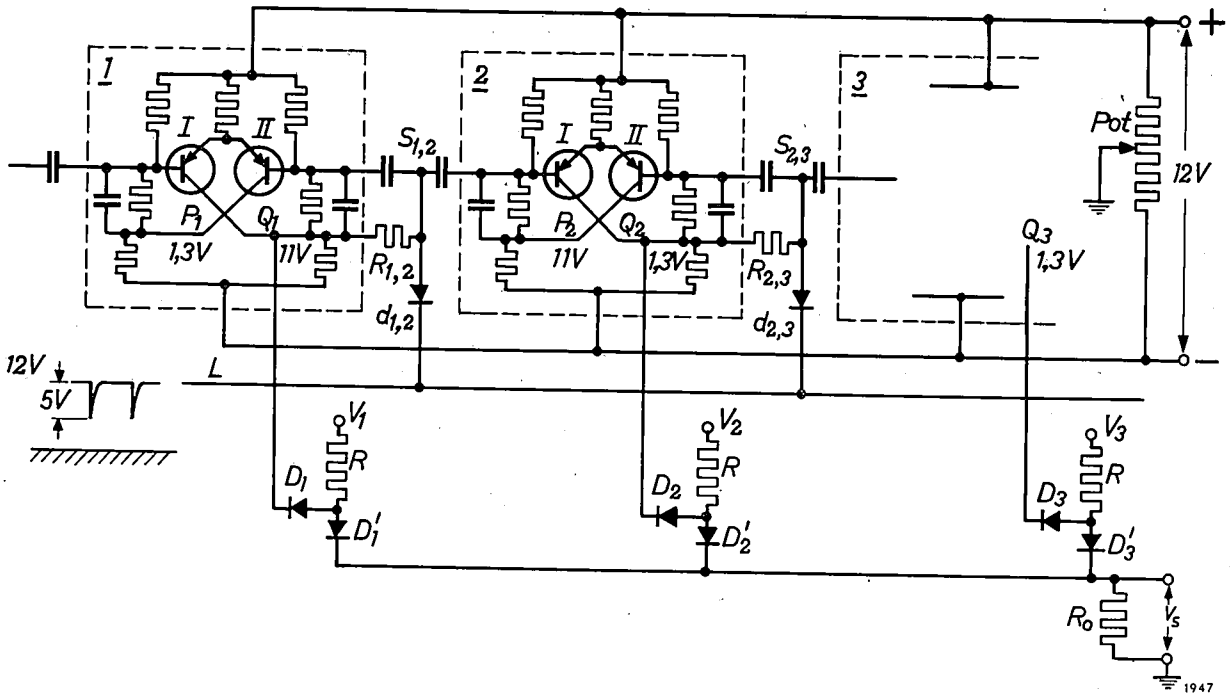


Fig. 7. The diodes D_1, D_2, D_3, ... are driven by 79 flip-flop circuits *1*, *2*, *3*, ... Flip-flop *1* is here in the "activated" state (transistor *I* conducting, *II* cut off, Q_1 is at the "high" potential, D_1 is cut off, $v_s = \frac{1}{2}V_1$). The other flip-flop circuits are in the quiescent state (transistor *I* cut-off, *II* conducting, $Q_2, \ldots Q_{79}$ are at the "low" potential, $D_2, \ldots D_{79}$ are all conducting, so that $V_2, \ldots V_{79}$ make no contribution to v_s). The successive flip-flop circuits are consecutively triggered by negative pulses on the line L.

Each flip-flop is provided with two transistors, type OC 71. There is always one flip-flop in the "activated" state whilst the other 78 are in the "quiescent" state. "Activated" means here that transistor *I* is conducting and transistor *II* non-conducting; point *P* thereby has a potential of 1.3 V and point *Q* a potential of 11 V in relation to the negative line. In the quiescent state the reverse is the case: transistor *I* is non-conducting, *II* conducting, *P* is at 11 V and *Q* at 1.3 V with respect to the negative line.

Points $Q_1, Q_2, \ldots Q_{79}$ are the corresponding points of the diode circuits (cf. fig. 5). Thus, only that diode circuit is operative whose associated flip-flop circuit is in the activated state; the voltages indicated in fig. 7 are for flip-flop *1* in the activated state.

The potentiometer *Pot*, the sliding contact of which is earthed, is adjusted so that the "low" voltage of all points *Q* with respect to earth has just the value needed in order to suppress the residual values of the voltage v_s across R_0 (see above). The "high" voltage of Q_1 with respect to earth (switch *1* being activated) is then still high enough to block diode D_1 completely, v_s then being approximately equal to $\frac{1}{2}V_1$.

The points $S_{1,2}$, $S_{2,3}$ etc. between the successive flip-flop circuits are connected via diodes $d_{1,2}$, $d_{2,3}$ etc. to a line L. With respect to the negative pole this has a potential of 12 V, superimposed on which are negative triggering pulses of approx. 5 V having a frequency of 2000 c/s. The anode of $d_{1,2}$ is connected via a resistor $R_{1,2}$ to Q_1, that of $d_{2,3}$ via $R_{2,3}$ to Q_2, and so on. The triggering pulses therefore find all diodes blocked except $d_{1,2}$ (since only Q_1 has the "high" voltage of 11 V). Via $d_{1,2}$ and the two coupling capacitors there thus arrives a negative pulse on the base of transistor *II* in flip-flop *1* and on the base of transistor *I* in flip-flop *2*. Both these circuits therefore change their state: *1* goes over into the quiescent state and *2* into the activated state. Upon the arrival of the next triggering pulse the same happens to *2* and *3*, and so on. In this way, the activated state passes $2000/79 \approx 25$ times-per second around the ring of flip-flop circuits and successively makes each of the 79 diode circuits momentarily conductive.

When flip-flop *2*, for example, changes to the activated state (fig. 7), the voltage of Q_2 jumps from 1.3 V to 11 V. The latter value is reached before the triggering pulse has ended, and there is therefore a danger that diode $d_{2,3}$ will open and cause the premature activation of flip-flop *3*. To prevent this happening, measures are taken to cause the anode voltage of $d_{2,3}$ to follow the voltage jump of Q_2 so slowly that $d_{2,3}$ remains blocked during the pulse. This means that the voltage variation on the anode of $d_{2,3}$ must have a sufficiently large time constant. (This time constant is determined primarily by the resistance $R_{2,3}$ and the capacitance of the coupling capacitors.) This applies, of course, to all 79 coupling networks.

The triggering pulses are derived by differentiation from a square-wave voltage generated by a multivibrator, using a double triode, type E 88 CC. Between the multivibrator and the line L a transistor type OC 77 is connected as an emitter-follower for matching between the impedances.

Replacement of the rotary switch S_2

A pulsed sweep voltage, like that obtained from the rotary switch S_2 (fig. 1) for the horizontal deflection (fig. 2*c*), is not easy to produce by electronic means. Other waveforms can be used, however, such as a staircase or a sawtooth waveform. A staircase voltage is again difficult to produce electronically [2]), whereas a sawtooth voltage is fairly simple to generate. *Fig. 8a* shows the appearance of

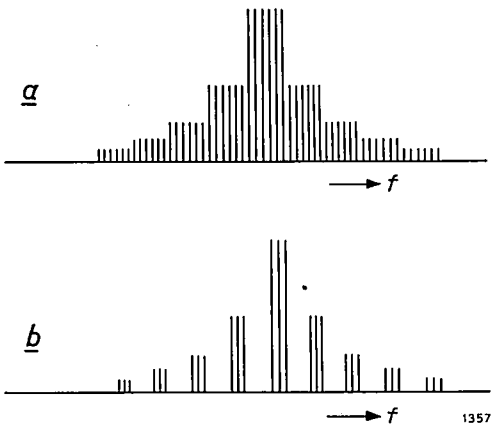


Fig. 8. *a*) Form of the spectrum of a sinusoidal voltage when the horizontal deflection is produced by a sawtooth voltage. *b*) The same, but now with the electron beam of the C.R.T. periodically suppressed for $\frac{1}{4}$ ms each $\frac{1}{2}$ ms by means of a square-wave voltage of 2000 c/s applied to the control grid of the oscilloscope. Each "bar" is in fact a sine wave train of about 12 cycles of the 50 kc/s voltage, but appears to the eye as a single vertical line.

a spectrogram of a sinusoidal signal when a sawtooth voltage is used for the horizontal deflection. A spectrogram of this kind is just as useful as the line spectrum obtained with the original apparatus (fig. 2*d*), but it is somewhat less suggestive. For this reason an attempt has been made to approximate to the latter spectrum in the following way. The oscilloscope tube is kept just below cut-off by a negative bias on the control grid, but the beam suppression is lifted once in every half cycle by a square-wave strobe voltage of 2000 c/s applied to the same electrode; this strobe voltage is derived from the multivibrator mentioned above. During every half millisecond, therefore, the electron beam is present for only $\frac{1}{4}$ millisecond. The effect of this on the spectrogram of a sinusoidal signal is shown in fig. 8*b*. The individual "bars" are not sharp lines as in fig. 2*d*, but sinusoidal wave trains of about 12 cycles of the voltage from generator G (fig. 1), which has a frequency of about 50 kc/s; the difference, however, is scarcely perceptible in practice.

The sawtooth voltage is generated by a Miller integrator based on a transitron-connected pentode (EF 80) [3]). Free-running the circuit has a frequency of approx. 20 c/s. When the 79th flip-flop circuit returns from the activated to the quiescent state, it produces pulses with a repetition frequency of $2000/79 \approx 25$ per second; these pulses are used to synchronize the transitron.

It is necessary to make the flyback time as short as possible since the information delivered by the filter circuits that come into operation during the flyback is lost. With this in view the capacitor in the sawtooth generator is made to charge up extremely rapidly through a cathode follower (one half of a double triode E 88 CC), which is driven by the transitron-Miller. In this way the flyback time has been reduced to less than 0.25 ms, i.e. less than 0.6% of the sawtooth period. (In a normal sawtooth generator the flyback time is generally about 10%.)

The short flyback time gave rise to a difficulty which, however, it proved possible to overcome completely. The difficulty is most evident if the input signal is a sinusoidal voltage with a frequency of 7440 c/s (the resonance frequency of the 79th filter network). The spectrogram then appears as illustrated in *fig. 9*. The deflections on the left of the spectrum create the false impression that the input signal also contains low-frequency components. This is particularly troublesome if the input signal really possesses a complicated spectrum.

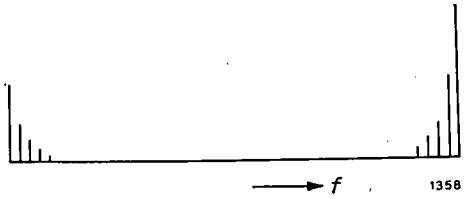


Fig. 9. Spectrogram of a sinusoidal voltage of frequency 7440 c/s. The deflections on the left suggest the presence of low-frequency components. They are due to the stray capacitance in parallel with R_0 being unable to discharge rapidly enough (cf. fig. 10).

[2]) See e.g. Philips tech. Rev. **12**, 288, 1950/51.

[3]) A transitron circuit is one of the pentode circuits with which a negative resistance is obtained; see for example F. E. Terman, Radio Engineer's Handbook, McGraw-Hill, London 1950, p. 318. A Miller integrator based on a transitron (see e.g. F. Kerkhof and W. Werner, Television, Philips Technical Library, 1952, p. 138 *et seq.*) produces a sawtooth generator that gives a very linear sawtooth wave-form.

The effect is due to the presence of an appreciable stray capacitance C_p in parallel with the resistor R_o across which the signal voltage v_s for the modulator appears. (This stray capacitance consists mainly of wiring capacitance and the total capacitance of the diodes connected to R_o.) The measured value of C_p was about 1500 pF, and with $R_o = 220$ kΩ this gives a time constant R_oC_p of 0.33 ms. In the flyback time of about 0.25 ms the voltage across C_p therefore drops only in the ratio of $1 : e^{-0.25/0.33} = 1 : 0.47$, and the deflection at the extreme left will consequently still be almost half as large as the deflection at the extreme right.

In principle the unwanted deflection on the left can be reduced in the following ways:

1) By reducing C_p; this, however, proved to be impracticable.
2) By reducing R_o, though at the expense of sensitivity.
3) By lengthening the flyback time, which would mean losing part of the spectrum; to remedy this, the ring would have to be extended by some additional flip-flop circuits.
4) By making C_p discharge rapidly during the flyback with the aid of some kind of switch.

The latter method proved to be a satisfactory solution. The switch used is a transistor (*fig. 10*) which is normally in the cut-off state but is made to conduct momentarily by the synchronizing pulse to the sawtooth generator during every flyback; this pulse originates from the 79th flip-flop circuit. The polarity of the voltages makes it necessary to use an *N-P-N* transistor, type OC 139.

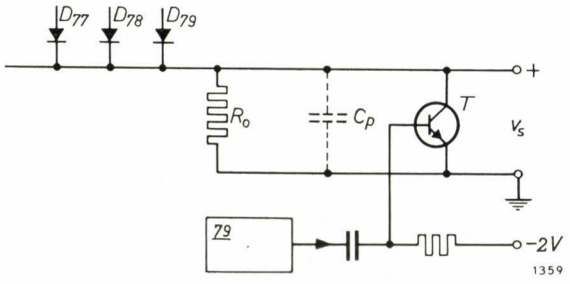


Fig. 10. The stray capacitance C_p (approx. 1500 pF) in parallel with the resistor R_o (220 kΩ) is rapidly discharged, during the flyback, through the transistor T (*N-P-N* transistor, type OC 139) which is periodically made to conduct by synchronizing pulses applied to its base. The synchronizing pulses are taken from the 79th flip-flop circuit. This arrangement suppresses the effect shown in fig. 9. (The diodes should be marked D_{77}', D_{78}', D_{79}'.)

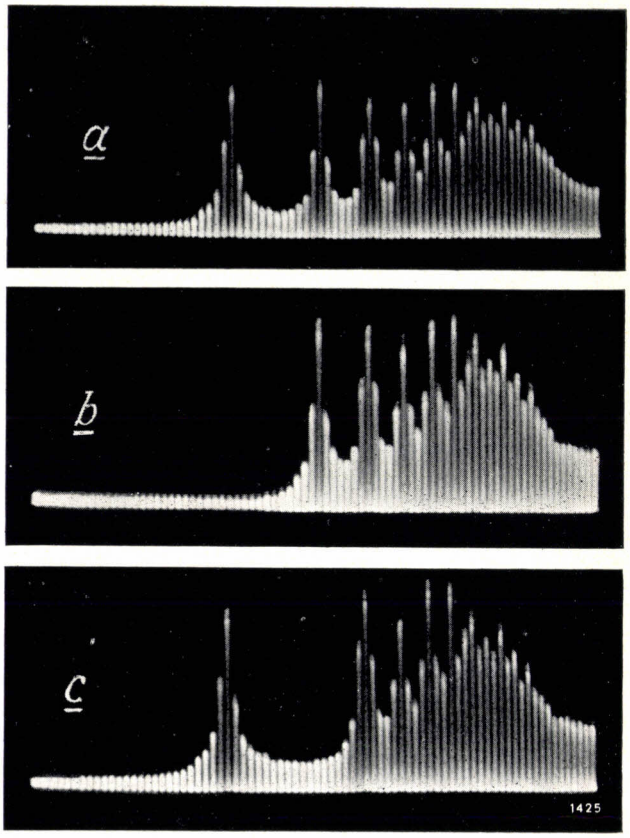
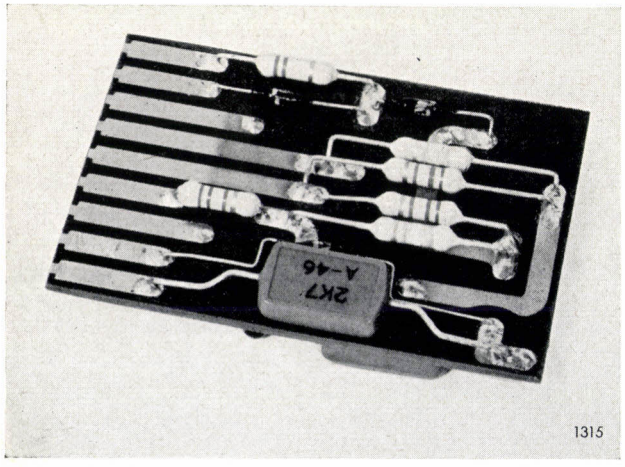


Fig. 11. Examples of spectrograms recorded on the acoustic spectrum analyser described in this article.
a) Fourier spectrum of a pulse.
b) Idem, with suppressed fundamental component.
c) As (*a*), with second harmonic suppressed.



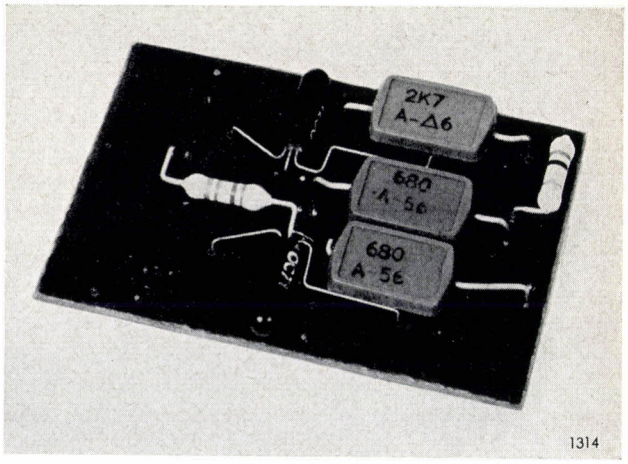


Fig. 12. The two sides of a printed-wiring board (42 × 76 mm), containing a diode switch circuit (D, D' and R), the associated flip-flop circuit and the components for coupling to the next stage. The board is held by spring contact clips, which serve as electrical connections to the rest of the circuit.

Photographs of a number of spectrograms are shown in *fig. 11a, b* and *c*. It will be noticed that (cf. fig. 2*d*) the negative deflections are now suppressed. This has the advantage of allowing the horizontal axis to be lower on the oscilloscope screen, enabling the positive deflections to be made larger.

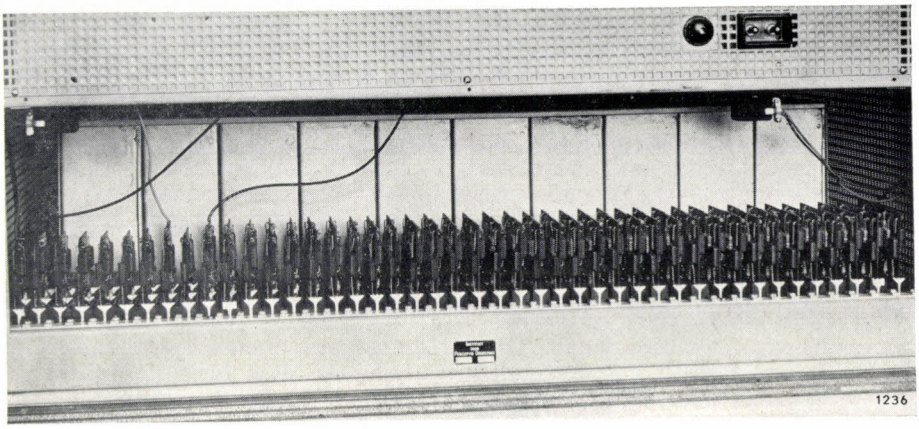


Fig. 13. Rear view of the interior of the acoustic spectrum analyser, showing the printed-wiring boards, one of which appears in fig. 12.

It should be mentioned in conclusion that the full width of the spectrum need not always extend from 86.5 to 7440 c/s; a smaller frequency band can be covered, which can then be displayed over the full width of the screen.

Constructional features

In many circuits of the acoustic spectrum analyser use is made of printed wiring [4]). For example, the 79 diode circuits with their flip-flop circuits are each assembled on a printed-wiring board 42×76 mm (*fig. 12*). The boards are held in place by spring clips which also constitute the electrical contacts. In this way the individual circuits are easily removable for testing and change where necessary. *Fig. 13* shows the boards in position in the apparatus, showing the neat, compact and accessible assembly.

Various pilot lamps are fitted to facilitate the localization of faults.

[4]) See e.g. Philips tech. Rev. **20**, 113, 1958/59.

Summary. Beljers's acoustic spectrum analyser contains 79 filters whose pass bands, each a semi-tone in width, consecutively cover a frequency range from 86.5 to 7440 c/s. In the original apparatus (of 1942) a rotary switch was used for causing each of the filter voltages to produce successive vertical deflections on an oscilloscope, and a second rotary switch was used to make the horizontal deflection increase stepwise logarithmically with the frequency. In the Institute for Perception Research at Eindhoven the apparatus has been modernized, one of the improvements being the replacement of both switches by silently-operating electronic switches using semiconductor diodes and transistors. Use is also made of printed wiring.

A SIMPLE METHOD OF DETERMINING THE THERMAL CONDUCTIVITY OF SOLIDS

by J. SCHRÖDER *).

536.21.08

The remarkable feature of the method described is that it involves no temperature or quantitative heat measurements. The procedure, using a small cylindrical sample, consists merely in taking a stop-watch reading (of a few minutes) and finding a value on a calibration chart.

Unlike many other material properties, the thermal conductivity has been measured on only relatively few solid materials. Moreover, the values reported for one and the same material often show considerable discrepancies, sometimes 50% or more. These discrepancies are partly explained by differences in the composition, degree of purity, pretreatment, etc., of the samples measured. They stem largely, however, from the difficulties involved in measuring this property reliably and reproducibly. These difficulties may also explain why relatively so few solids have been subjected to thermal conductivity measurements, in spite of the growing importance of this property for scientific as well as technical reasons. It is of interest scientifically because no comprehensive theory has yet been put forward to explain the mechanism of (non-metallic) thermal conduction in solids. It is of technical importance in connection with all solid-state applications involving the conversion of thermal energy. In some applications, such as Peltier cooling and thermo-electric energy production, heat conduction in fact governs the efficiency that can be achieved.

The methods of determining thermal conductivity, λ, described in the literature, are too numerous to summarize here [1]). Common to all of them is the measurement of the temperature difference ΔT across a sample of length l — conveniently done with thermocouples — and of the quantity of heat Q flowing through the sample in a given time t. The thermal resistance R is then given by

$$\frac{1}{R} = \frac{A}{l}\lambda = \frac{Q}{t\Delta T}, \quad \ldots \quad (1)$$

where A is the cross-section of the sample. The value of Q is usually derived from the electric power used for the heat supply, but it can also be measured calorimetrically. The most accurate values of λ are obtained by measurements in the steady state, to which (1) applies. This involves a lengthy measurement, however, since it is necessary to wait until this state is reached, and particularly since, to minimize heat losses, the sample and the heat source must be shielded and the shielding must also be brought to the requisite steady temperature. Moreover, these precautions entail the use of rather cumbersome equipment. In many methods, too, a good deal of time is spent on preparing the sample, so that all in all no more than one sample a day can be investigated.

In the following a new method is described, which uses simple equipment and makes it possible to measure the thermal conductivity of solids at room temperature quickly and accurately. An incidental advantage of the method is that the samples required are not only easy to make but can also be fairly small — a most important requirement where many solids are concerned and one which is certainly not met by certain older methods. The same method, with some supplementary equipment, can also be used for determining λ values at temperatures well above and below ambient, viz. between −200 and +400 °C.

The basis of the new method is the maintenance of a fixed temperature difference between the ends of the sample, the temperature of each end being held constant by contact with two boiling liquids. What is measured is the time in which, in the steady state, a certain quantity of heat passes through the sample; this quantity is established quite simple by the quantity of liquid made to evaporate at the "cold" end of the sample.

The principle of the method is illustrated in *fig. 1* [2]). In the lower vessel A a pure liquid a is brought to the boil. The vapour flows over the silver plate S_1, is condensed in the cooler K_1, and the condensate returns through the overflow pipe L into the vessel. In this way, plate S_1 is very effectively kept at a constant temperature, the boiling

*) Zentrallaboratorium Allgemeine Deutsche Philips Industrie GmbH, Aachen Laboratory.

[1]) See e.g. M. Jakob, Heat transfer, Wiley, New York 1949, Part 1, Chapter 9. Also Archiv für technisches Messen V 9213 1 to 4.

[2]) The apparatus will shortly be put on the market by Colora Messtechnik GmbH, Lorch/Würtemberg, Germany.

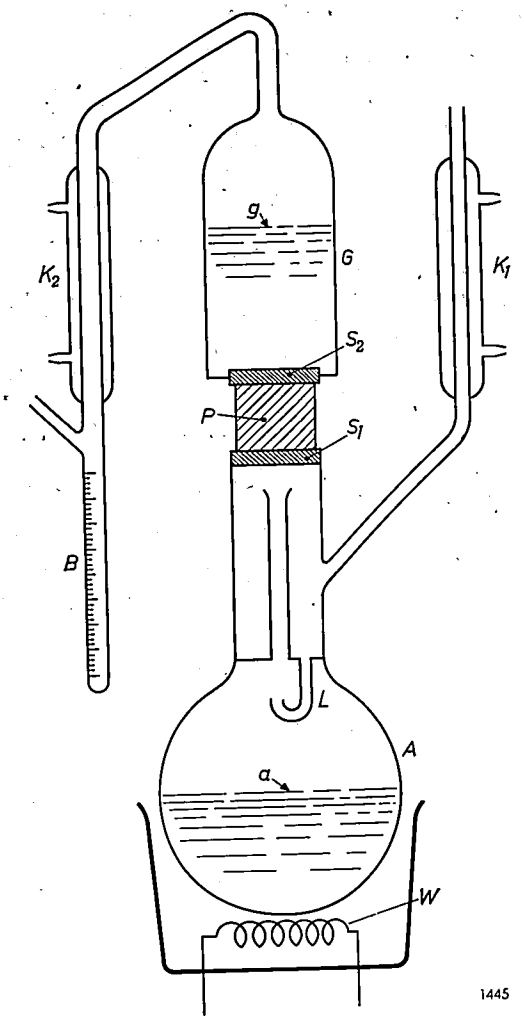


Fig. 1. Principle of the new method of determining the thermal conductivity of solids. P sample of material under investigation. A and G vessels containing two different liquids (a and g). W heating element that keeps a on the boil. S_2 and S_1 silver plates, ensuring good thermal contact of P with the liquid g and the vapour of a, respectively. K_1 condenser and L overflow pipe for returning the condensate of a. The liquid g, whose boiling point is roughly 10 °C lower than that of a, is brought to the boil by the heat transmitted through the sample P. K_2 condenser and B tube, graduated in millilitres, for collecting and indicating the vaporated quantity of liquid g.

point T_a of the liquid, for if S_1 drops only slightly below that temperature, vapour condenses on the plate and transfers to it its heat of condensation, causing the temperature of the plate to rise again to T_a. The upper vessel G contains another pure liquid, g, whose boiling point is, say, 10 °C lower than that of liquid a. At the bottom of this vessel there is another silver plate, S_2. Fitted between the two plates, S_1 and S_2, is a cylindrical sample P of the material under investigation. The heat that flows through this sample from S_1 to S_2 brings the liquid g to the boil, thereby keeping the temperature of S_2 constant at the boiling point T_g of g. Between the two silver plates there is therefore a constant temperature difference $T_a - T_g$. The vapour from liquid g is condensed in the cooler K_2 and the condensate is collected in a tube B, marked with a millilitre scale.

As soon as the steady state of heat flow is reached — which is reached when about 0.1 ml of condensate has collected in B — the time t taken for a certain liquid volume V, say 1 ml, to flow into B is determined with a stop watch. Let S be the heat of vaporization of the liquid g per unit volume, then eq. (1) gives:

$$\lambda = \frac{l}{A}\frac{1}{R} = \frac{l}{A}\frac{VS}{t(T_a - T_g)}. \quad \ldots \quad (2)$$

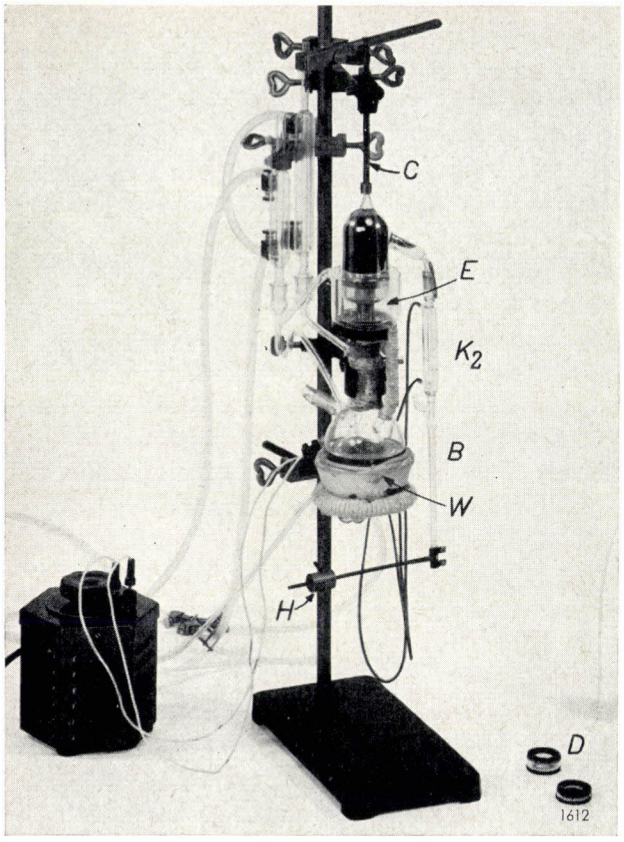
Given the data on the liquids and the dimensions l and A of the sample we can calculate λ from the measured values of V and t.

The method shows a remarkably high degree of accuracy, owing to the fact that the effect of heat losses between S_1 and the liquid g is kept relatively very low. This is done primarily by choosing the dimensions of the sample so as to ensure a reasonably large heat flow through the sample. For instance, if the substance is a poor conductor of heat, the sample taken will be in the form of a thin disc. For a better conductor, assuming the same temperature difference, a thicker sample will be taken, for an unduly large heat flow would cause the liquid g to boil too turbulently, with the risk of liquid splashing over into tube B. Furthermore, if the appropriate quantity of liquid distils too quickly into B, the time measurement is less accurate. The time t should preferably be between 100 and 1000 seconds, and this can be arranged not only by a suitable choice of the thermal resistance of the sample but also by using suitable liquids (fixing the values of S and T_a and T_g).

Heat may be lost from the upper vessel as well as from the sample, causing errors in V through insufficient evaporation or premature condensation. These errors can be minimized by enclosing the upper vessel G in a vacuum jacket (dewar flask); see the photographs of the set-up in *fig. 2*.

The method as described enables absolute determinations of λ to be made. If a number of calibrated samples are available, λ can be determined even more easily by a comparative measurement [3]). The times t found for the calibrated samples are plotted in a diagram against the known values of the thermal resistance R. On the calibration curve thus

[3]) The standard samples were prepared for us by K. H. Bode and W. Fritz of the Physikalisch-Technische Bundesanstalt, Brunswick. The values of thermal resistance determined by them were accurate to within ± 1% (Z. angew. Phys. **10**, 470-479, 1958). We take this opportunity to thank them for their valuable assistance.



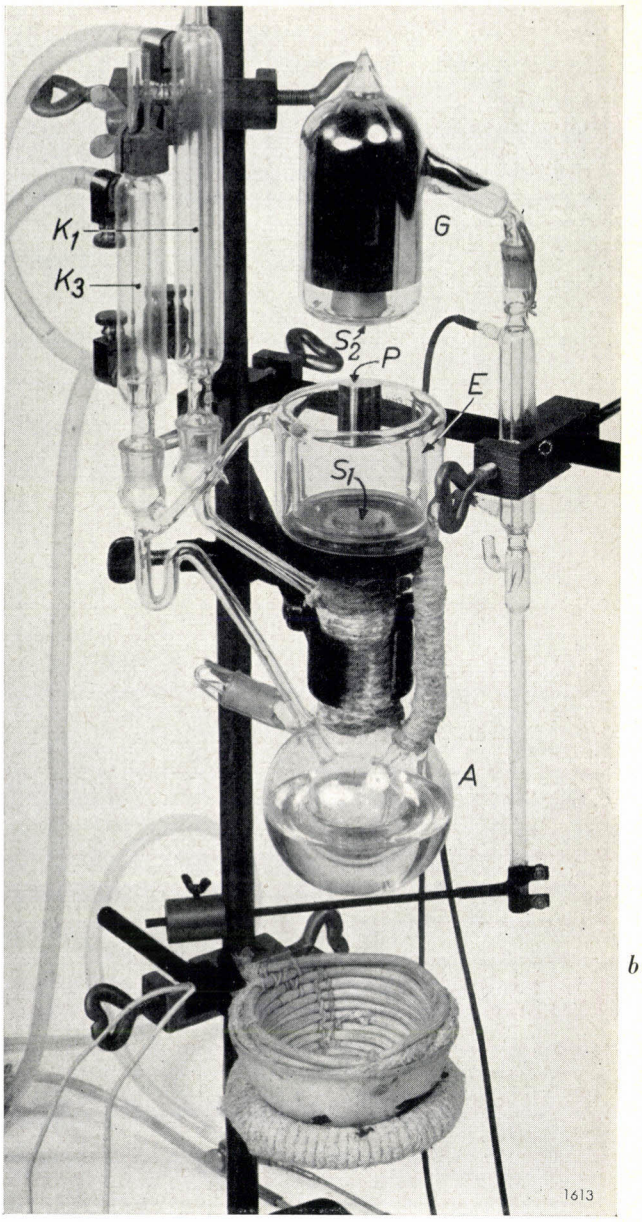


Fig. 2. The apparatus, *a*) ready for use, *b*) partially dismantled (for this photograph the sample P was suspended by a thread between the two silver plates S_1 and S_2).

The upper vessel G is in the form of a dewar flask, with the sealed-off pinch at the top and the condenser K_2 and graduated tube B at the side. It is placed over the sample and pressed down by the long helical spring C. K_2 and B are balanced by an arm with the weight H. One of the rings D is placed around the sample to give thermal insulation. A refinement not shown in the arrangement in fig. 1 is a double-walled glass vessel E surrounding the sample; between the walls the vapour can be circulated of a liquid whose boiling point is between those of a and g. The vessel containing this liquid can be seen inside vessel A, and associated with it is a condenser K_3 with return pipe.

obtained we can then, after having measured the time t, read off directly the value of R for a given sample, and from this value, together with the dimensions of the sample, we can calculate the value of λ.

It should be noted that for these relative measurements it is not necessary to know the exact values of V, T_a and T_g. This at once eliminates the errors that can creep into a calculated λ value if *a*) the boiling points T_a and T_g should differ slightly from the assumed values owing to impurity of the liquids, and if *b*) the ends of the sample should not attain the exact temperatures T_a and T_g owing to contact resistances to the heat transfer between the silver plates and the sample. It should of course be ensured that these resistances are equal for all measurements, in order to obtain reproducible results. The same applies to the magnitude of the small residual heat loss, to which we shall return presently.

The calibration curve, as can be seen in *fig. 3*, is very closely a straight line. The negligible spread of the points is evidence of the high degree of reproducibility of which the method is capable. According to eq. (1) the curve should be a straight line through the origin. In the diagram this is not so, the line cutting off a section R_0 on the ordinate axis. This corresponds to the thermal resistance of the various heat-transfer surfaces. For each pair of liquids used the calibration line must be plotted afresh, but since it is a straight line it is sufficient to measure two standard samples, possibly with a third for verification.

By this comparative method, a measurement, including the preparation, lasts no more than 5 to 15

minutes. The error of measurement is $\pm$ 3% at the most. No temperatures need be measured, and it is not necessary to know the exact values of the boiling point and heat of vaporization of the liquids used. The liquids need not even be pure, provided the same liquids are used for calibration and measurement. The ease and speed with which a sample can be measured by this method open up the practical possibility of investigating systematically large series of specimens, as for example mixed crystals of varying composition.

Some particulars follow concerning the equipment and the samples.

Various measures are taken to ensure that the temperatures of the silver plates S_1 and S_2 will be accurately reproducible. To avoid any delayed-boiling effect in liquid g, capillary holes are made in the plate S_2. To increase the surface area for the heat transfer, both plates S_1 and S_2 are provided with grooves on the side in contact with the vapour and liquid, respectively. At the other side the plates are ground flat to obtain the best possible contact between sample and plates; also, the sample length is thereby precisely defined. To reduce the contact resistances a heat-transfer fluid is applied between the faces, and the plates are held against the sample under considerable (and constant) spring pressure (see fig. 2a). In this way the variations in the thermal resistance across the heat-transfer surfaces are reduced to such an extent that their effect in all measurements remains below the limits of accuracy determined by other factors.

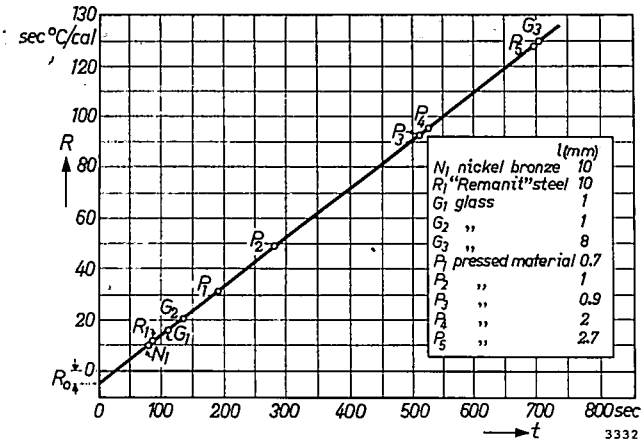


Fig. 3. Calibration chart for an apparatus using carbon disulphide and ether as boiling liquids (boiling points 46.3 and 34.5 °C, respectively). For ten calibrated samples (see inset) for which the thermal resistances R were known — values differing by a factor 10 for these samples — R is plotted against the measured time t taken for 1 ml of condensate to collect in tube B. The calibration points lie on a very good straight line, demonstrating the accuracy of the method.

Extrapolation of the line to $t = 0$ gives the correction R_0 to be applied in an absolute measurement of thermal resistance (R_0 takes account of contact resistances as well as heat losses).

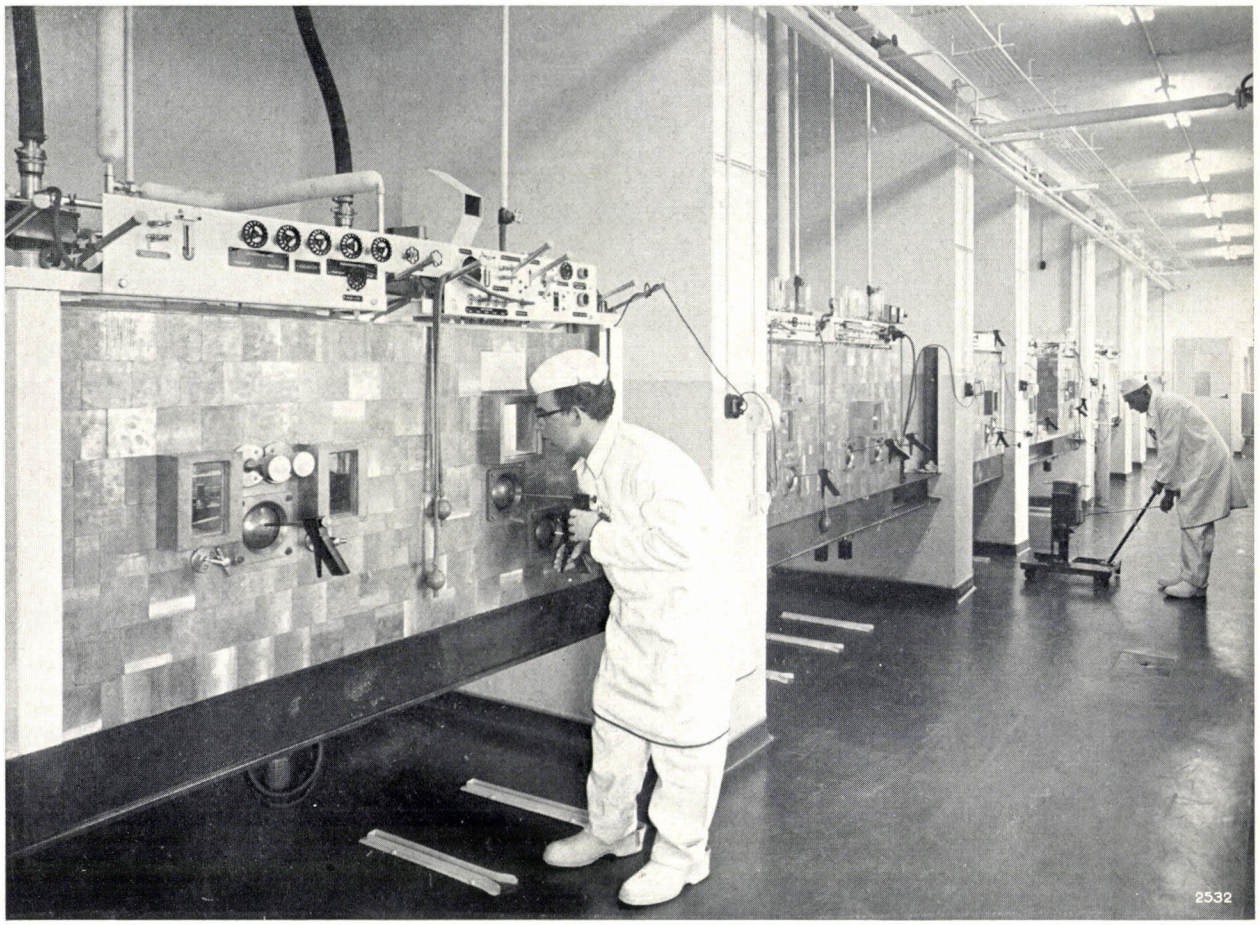
It has already been mentioned that the upper vessel G is enclosed in a dewar flask to minimize heat losses. The sample must also, of course, be provided with heat insulation; a series of rings are available to surround the sample (two such rings can be seen at the right of the stand in fig. 2a). In order to carry out thermal conductivity measurements at temperatures very different from room temperature, both vessels and the sample must be surrounded by shields, held at the appropriate temperatures by baths in which the same liquids as in A and G are kept at the boil. For measurements below room temperature, where G and A thus contain liquids that are gaseous at room temperature, the vapour from G need not be condensed in a cooler but can be collected and measured in a gas meter.

As regards the liquids themselves, pairs can be found for the whole temperature range from —200 to +400 °C having suitable boiling points differing by 5 to 15 °C. If necessary, the same liquid can be used for both vessels, and the boiling point in vessel A raised slightly by a small and constant increase of pressure, controlled by a pressurestat.

The samples used for both calibration and measurement are cylindrical in form, ranging in diameter from 16 to 18 mm and in length from 0.5 to 30 mm, depending on the thermal conductivity of the material. Samples of material of high thermal conductivity (pure metals) can be made in the form of hollow cylinders, say 10 mm in length and with a wall thickness preferably not less than 1.5 mm.

Summary. The two ends of a cylindrical sample, about 18 mm in diameter and 0.5-30 mm in length, are kept at constant temperatures, by contact with two boiling liquids of suitable boiling points, viz. differing by about 10 °C. The *time* is measured in which a quantity of heat flows, in the steady state, through the sample. This quantity is fixed very simply by the evaporation of a fixed quantity of liquid at the "cold" end, collected as condensate. In practice, two calibrated samples are measured first and a calibration line is plotted from which the thermal resistance of subsequent samples can be found at a glance. It is then not necessary to know the exact boiling points of the two liquids, nor the absolute amount of evaporated liquid, nor its heat of vaporization. A measurement takes 5 to 15 minutes and the error is not more than $\pm$ 3%. With some additional equipment the method can also be used for measurements at other than room temperature, between —200 and +400 °C.

CHEMICAL FORMULATION OF RADIO-ISOTOPES



The photograph shows a part of the Isotope Laboratory of N.V. Philips Duphar in Amsterdam, where radio-isotopes produced in the Amsterdam cyclotron or in reactors at Kjeller (Norway), Mol (Belgium) or Saclay (France) are chemically processed. Among the isotopes handled are ^{131}I, ^{198}Au, ^{206}Bi, etc. Workers in the laboratory are protected from radiation hazards by performing all operations behind a thick lead wall, in fume cupboards with an underpressure of several cm water. The latter measure is to prevent contamination of the laboratory atmosphere by radioactive dust. A further safety measure is that all workers in the laboratory wear special clothing.

Manipulations of the radioactive substances are carried out with the aid of the usual remotely-controlled tongs. The man in the foreground is operating such a device. Windows of lead glass 20 cm thick make it possible to direct the tongs and follow the process. On the panel above the lead wall are cocks admitting water, steam, air, vacuum and certain standard reagents to the reaction vessel. Of the projecting rods on the panel, some serve for the opening and closing of the passage connecting one fume cupboard with the next, while others are for the operation of a balance for weighing the radioactive materials. The balance scale is visible on a ground-glass plate contained in the box seen just above the fume cupboard.

The fume cupboard at which the operator is busy is a pipetting cupboard, for transferring measured quantities of radioactive solutions. On the left of the operator are two rubber squeezers for sucking the solutions into the pipettes. In the cupboard on the left preparations for medical use can be sterilized. The two knobs between the two lead-glass windows open and close the autoclaves.

The fume cupboards are accessible only from underneath through special doors. The operator in the background is moving a transportable vessel containing the radioactive raw material to a position under the door of a fume cupboard. The guides in the floor ensure that the vessel is placed precisely under the door. The vessel is then raised until it just touches the door; the latter is opened from outside, whereupon the vessel is raised further into the cupboard. The contents may now be manipulated in the fume cupboard. The same procedure in the reverse order accomplishes removal of the vessel and closing of the door.

PERFORMANCE TESTS ON LOUDSPEAKERS

by M. T. HAITJEMA, W. KOPINGA and S. J. PORTE. 53.08:621.395.623.7

The article below describes a number of established methods of loudspeaker testing used over the years at Philips. These methods have proved particularly valuable in the design of loudspeakers and loudspeaker cabinets. Some are also well adapted for sampling inspection during manufacture.

The loudspeaker tests discussed in this article are concerned with:

a) the variation of sound pressure as a function of frequency,
b) the variation of loudspeaker impedance as a function of frequency,
c) the resonance frequency,
d) transients, such as occur when the signal applied to the loudspeaker suddenly drops to zero,
e) directivity patterns, giving the sound pressure in different directions at constant frequency in polar coordinates, and
f) the variation of loudspeaker efficiency as a function of frequency.

Sound pressure, impedance and efficiency are recorded as a function of frequency on a strip-chart recorder; the transients and the directivity patterns are displayed on a cathode-ray oscilloscope [1]).

Sound pressure as a function of frequency

The procedure for recording the variation of the sound pressure of a loudspeaker as a function of frequency (frequency response) is as follows. The signal from a signal generator is applied to the loudspeaker under test via an amplifier. Mounted in line with the axis of the loudspeaker [2]) is a condenser microphone, the signal from which is amplified and recorded. Loudspeaker and microphone are situated in an acoustically dead (non-reverberant) room at a distance of 50 cm apart (*fig. 1*). This distance proved to be the best compromise in connection with the imperfections of the dead room and the dimensions of the loudspeaker (provided that the diameter of the cone is not larger than about 30 cm).

A small motor drives a rotary capacitor in the signal generator such that the frequency is swept from 0 to 20000 c/s, the range from 0 to 100 c/s linearly with time and the range from 100 to 20000 c/s logarithmically (equal times for each swept octave). The same motor drives the strip-chart recorder. To allow for possible small discrepancies between the frequency scale printed on the paper strip and the true frequency variation, the generator signal is automatically interrupted for a moment as its frequency passes 1000 c/s.

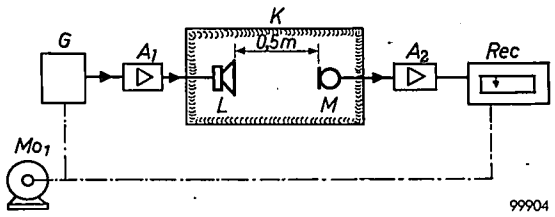


Fig. 1. Block diagram of apparatus for recording the sound pressure of a loudspeaker L as a function of the frequency of a signal generator G. A_1 amplifier with 50 W output stage in push-pull. M condenser microphone. A_2 microphone amplifier. *Rec* recorder. Mo_1 motor, which causes the frequency to sweep from 0 to 20000 c/s and also drives the strip-chart recorder. K dead room.

The amplifier A_1 between the signal generator and the loudspeaker has a 50 W output stage in push-pull fitted with four EL 34 pentodes in class AB. The output transformer is provided with taps for matching to loudspeaker coils of different impedance. Depending on whether the output current or output voltage is to be kept constant with varying frequency, 20-fold negative current or voltage feedback can be switched in. The frequency response is recorded at a constant current, viz. $\sqrt{50 \times 10^{-3}/R}$ amp, where R is the DC resistance of the loudspeaker coil in ohms. This means a power consumption of 50 mW at low frequencies.

The microphone amplifier A_2 is corrected so as to give the condenser microphone plus amplifier a flat frequency response from 20 to 20000 c/s.

The recorder is fitted in the usual way with an input potentiometer whose sliding contact is coupled with the stylus. The arrangement is such that the stylus is at rest only when the sliding contact is at

[1]) The methods of recording directivity patterns and efficiency curves were described briefly in T. Ned. Radiogenootschap **23**, 303-309, 1958 (No. 6).

[2]) All the tests described here were made on loudspeakers without baffles. Thus, the properties of the loudspeaker itself were measured, and not the additional properties of a baffle.

a point having a potential of 10 mV with respect to earth (the deflection is thus zero for input signals smaller than 10 mV). The relation between the deflection of the stylus and the input voltage depends on the resistance law of the potentiometer. In the present case a decibel scale is required for the ordinate of the response curve, whereas a linear scale is needed for the impedance characteristic, discussed below. Two interchangeable input potentiometers are therefore used, one of which gives a logarithmic and the other a linear scale, both for input voltages from 10 mV upwards.

Fig. 2 shows a loudspeaker frequency response curve recorded by the method described.

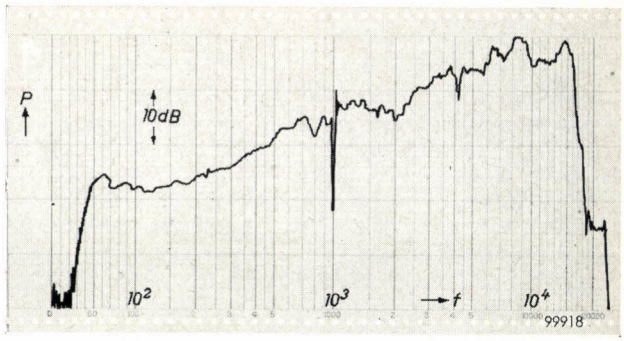


Fig. 2. Recording of sound pressure P as a function of frequency f (frequency response), obtained by the method illustrated in fig. 1. Sound-pressure scale logarithmic; frequency scale linear from 0 to 100 c/s, logarithmic from 100 to 20 000 c/s. At 1000 c/s there is a marker dip, obtained by momentarily interrupting the signal at this frequency.

Electrical impedance as a function of frequency

The electrical impedance of a loudspeaker is the ratio of the voltage across the loudspeaker coil to the current through it. When the current is kept constant, the voltage is thus proportional to the impedance. For recording the impedance/frequency curves the loudspeaker is again fed via the above-mentioned push-pull amplifier (with negative current feedback) with an audio signal from the signal generator, and the voltage across the loudspeaker coil is recorded. In this case the recorder uses the potentiometer giving a linear scale.

Fig. 3 shows two examples of impedance curves. Unlike curve *1*, curve *2* was recorded on a loudspeaker fitted with a short-circuiting ring [3]). This causes the impedance to rise much less sharply with frequency, thereby improving the reproduction of high notes.

Determination of the resonance frequency

The resonance frequency of a loudspeaker is the frequency at which the electrical impedance shows a maximum, by analogy with an inductance and capacitance in parallel which, in the region of the resonance frequency, form the electrical equivalent of the loudspeaker. The resonance frequency depends on the construction of the loudspeaker and also, where present, on the cabinet. The importance of the resonance frequency lies in the fact that it constitutes roughly the lower limit of the spectrum which the loudspeaker is capable of reproducing. (Below the resonance frequency the sound pressure drops with decreasing frequency by 18 dB per octave in the case of a loudspeaker without baffle, and by 12 dB per octave in the case of a loudspeaker in an infinitely large baffle.)

The resonance frequency is determined by keeping the voltage across the loudspeaker constant (negative voltage feedback) and by adjusting the frequency of the signal generator until the current is minimum; this frequency is the resonance frequency. The constant voltage is chosen such that the minimum current is 50 mA for loudspeakers of low resistance (of the order of 10 ohms), and 5 mA for loudspeakers of high resistance (400 or 800 ohms) [4]). At these currents the amplitude of the cone is so small that non-linearity in the deformation of the centering ring or of the rim of the cone can be neglected.

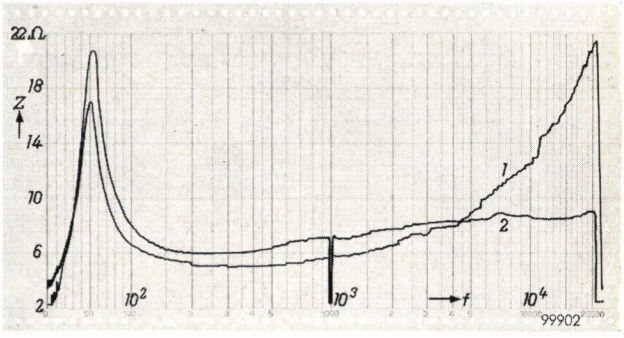


Fig. 3. Recordings of the impedance Z of a loudspeaker (linear scale) as a function of frequency f (in c/s). Curve *2* refers to a loudspeaker fitted with a short-circuiting ring [3]), curve 1 to a loudspeaker without such a ring. The flatter shape of curve *2* implies better reproduction of the high notes.

For determining the resonance frequency of loudspeakers in manufacture, an instrument giving a direct reading is preferred. An instrument of this kind has been devised, in which the loudspeaker under test functions as the LC circuit of an oscillator. The latter is provided with both positive and negative feedback. Both feedback circuits contain

[3]) See Philips tech. Rev. **18**, 313, 1956/57.

[4]) For high-resistance loudspeakers, see Philips tech. Rev. **19**, 42, 1957/58.

non-linear elements which are dependent on the current. This makes the current through the loudspeaker virtually independent of the damping at resonance, the value of which may differ considerably for loudspeakers of different types. The frequency at which the system oscillates is the resonance frequency of the loudspeaker and can be read from a direct-reading frequency meter.

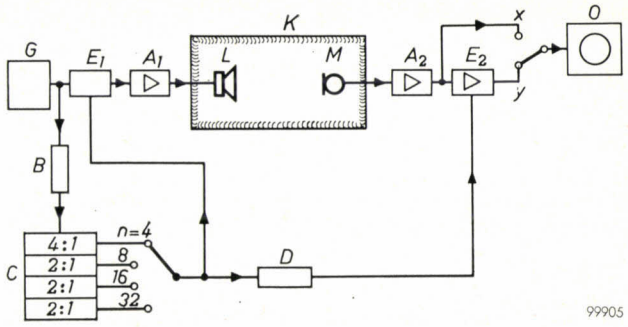


Fig. 4. Block diagram of the equipment for displaying transients. *G* signal generator. E_1 electronic switch producing wave trains from the sinusoidal voltage. A_1 push-pull amplifier. *L* loudspeaker under test and *M* microphone in dead room *K*. A_2 microphone amplifier. E_2 electronic switch for suppressing the microphone signal during the presence of the wave train. *O* cathode-ray oscilloscope. *B* circuit for converting the sinusoidal waveform into a square-wave form. *C* frequency dividers controlling electronic switches E_1 and E_2; they deliver square-wave forms of fundamental frequencies $^1/_4$, $^1/_8$, $^1/_{16}$ and $^1/_{32}$ times the frequency *f* of the signal generator. *D* device which delays the control signal for E_2 by an amount equal to the time taken by the sound to travel from *L* to *M*.

Oscilloscopic display of transients

Of great importance to the good reproduction of speech and music is the way in which the loudspeaker responds to plosives and the sudden entry of musical instruments, that is its response to abrupt initiation or removal of the electrical signal. The resultant transients, and in particular the subsidence transient following the interruption of the signal, can be studied by the following method.

The loudspeaker is supplied with a periodically interrupted sinusoidal signal (sine wave trains). The acoustic signal delivered by the loudspeaker, and picked up by a neighbouring microphone, is displayed on a cathode-ray oscilloscope. The oscillogram shows the initiation transient as well as the subsidence transient. The display of the latter can be improved by inserting a device between microphone and oscilloscope which suppresses the microphone signal for the duration of each applied wave train, so that only the transient proper remains visible. Loudspeaker and microphone must be set up in an acoustically dead room, as otherwise reverberation would appear as spurious deflections in the oscillogram.

Fig. 4 shows the block diagram of the circuit. An electronic switch E_1 between the signal generator *G* and the push-pull amplifier A_1 produces the sinusoidal signals by intermittently passing the voltage from the signal generator for a certain number of cycles *n* and blocking it for the same length of time; the signal is switched on and off at the moment the sine wave passes through zero. The electronic switch is controlled by a square-wave voltage supplied by the frequency-divider stages *C* (bistable multivibrators), one or more of which can be switched in to set *n* at 4, 8, 16 or 32 cycles. The first divider stage receives its input signal from the signal generator via the circuit *B*, which converts the sinusoidal signal into a square-wave voltage. *Fig. 5a* shows an oscillogram of the microphone signal for $f = 2500$ c/s, $n = 16$.

To suppress the microphone signal during each applied wave train, an electronic switch E_2 is placed between the microphone amplifier A_2 and the oscilloscope. This switch is controlled by the same square-wave voltage as E_1, but roughly in anti-phase and moreover with a fixed time-delay equal to the time which the sound from the loudspeaker takes to reach the microphone. The way in which the control signal is delayed is explained in *fig. 6*. The suppression is effective when the oscilloscope in fig. 4 is

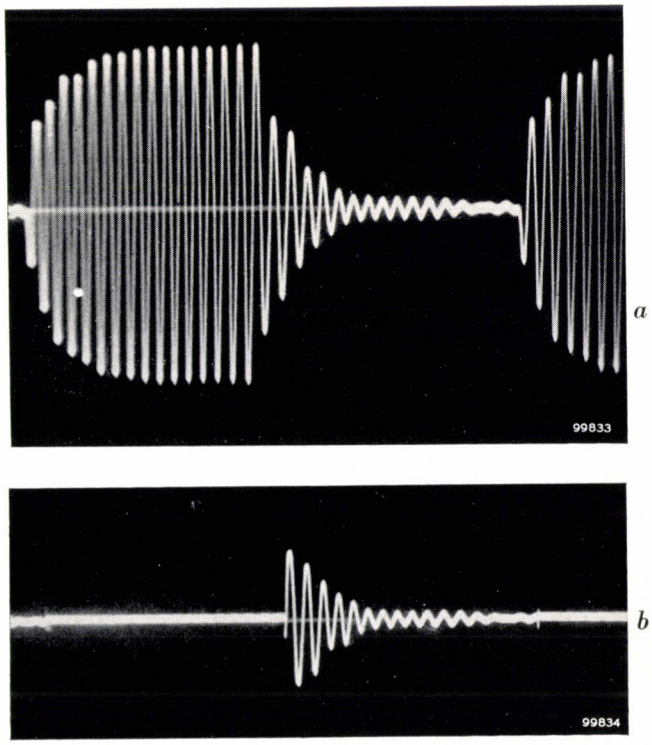


Fig. 5. *a*) Oscillogram of the microphone voltage in fig. 4 (oscilloscope connected to point *x*).
b) Oscilloscope connected to point *y* in fig. 4: the microphone signal is suppressed for the duration of the wave train itself, leaving only the transient visible.

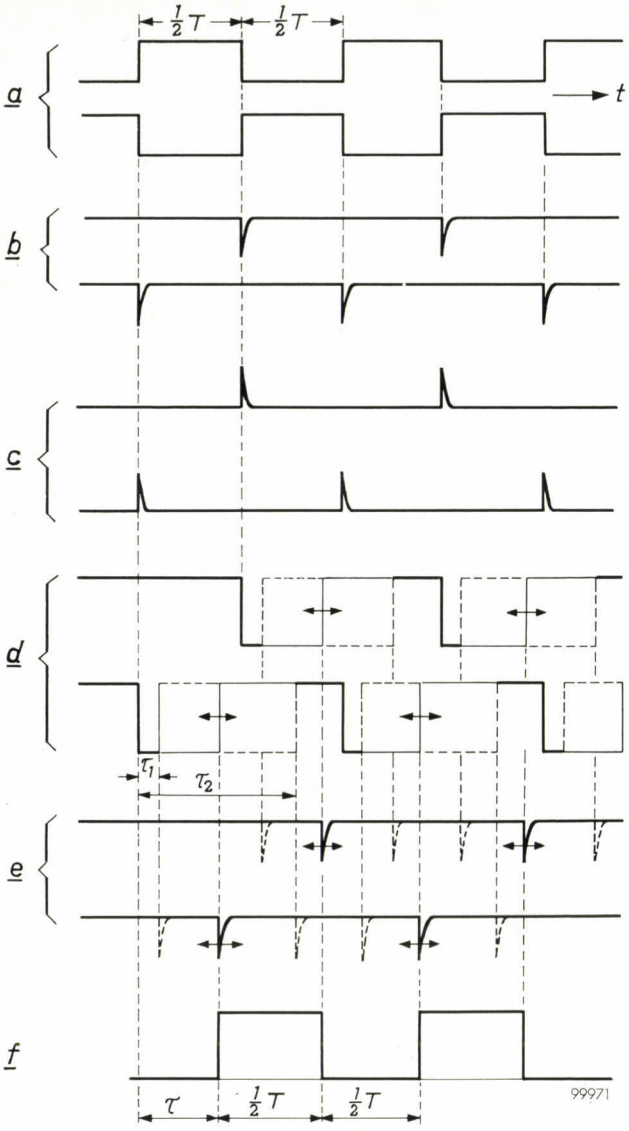


Fig. 6. Illustrating the way in which network D (fig. 4) delays the control signal for the electronic switch E_2 by a variable time τ, independent of the frequency f.
a) Square-wave voltages in anti-phase, derived from the frequency dividers C (fig. 4). A half period $\frac{1}{2}T$ comprises n cycles of the signal-generator voltage ($n = 4$, 8, 16 or 32).
b) Negative pulses obtained by differentiating the square-wave signals (*a*) and blocking the positive pulses.
c) Positive pulses obtained by amplifying the pulses in (*b*). The pulses (*c*) control a monostable flip-flop which, after each triggering pulse, returns to the quiescent state with an adjustable delay (by means of RC network with variable resistance).
d) Output of monostable flip-flop. The reset time of the flip-flop can be varied between the limits τ_1 and τ_2.
e) Pulses produced by differentiating, clipping and amplifying the signal in (*d*). These pulses can be varied in time between the limits τ_1 and τ_2, and are used to trigger a second multivibrator.
f) Square-wave voltage delivered by the second multivibrator. The signals in (*d*) are so adjusted that the square-wave voltage (*f*) has the same half period $\frac{1}{2}T$ as the signals in (*a*), but are displaced relative thereto by a variable interval τ. The signal (*f*) controls the electronic switch E_2 (fig. 4).

connected to point y. The resultant oscillogram of the transient alone is shown in fig. 5*b*.

As regards the transients at lower frequencies it makes a great difference whether the push-pull amplifier A_1 operates with negative current or voltage feedback. In the first case the amplifier behaves like a current source. Its internal resistance — in parallel with the loudspeaker — is then about 100 × higher than the loudspeaker resistance: the loudspeaker is thus only feebly damped and the transient takes a relatively long time to subside. With negative voltage feedback the amplifier behaves like a voltage source, the internal resistance is low (about 0.1 ohm): the loudspeaker is then strongly damped and therefore the transient is short-lived. The difference can be seen in the oscillograms in *fig. 7a* and *b*.

Some particulars of the oscilloscope are discussed at the end of the following section.

Display of directivity patterns

The directivity pattern of a loudspeaker presents the variation of the sound pressure — at a given distance and at constant frequency — in different directions. The form of the pattern depends on the frequency and on the dimensions and construction of the loudspeaker and of the cabinet or baffle, if present.

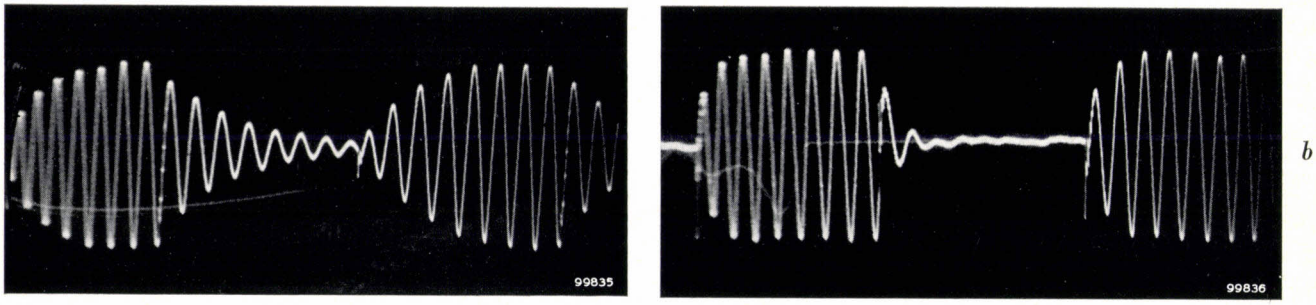


Fig. 7. Oscillograms of the microphone voltage (oscilloscope in fig. 4 connected to point x) at $f = 180$ c/s, $n = 8$. In (*a*) the amplifier A_1 operated with negative current feedback so that the loudspeaker was only weakly damped; in (*b*), the amplifier operated with negative voltage feedback so that the loudspeaker was heavily damped.

In our test set-up the loudspeaker is rotated in front of a fixed microphone in an acoustically dead room (*fig. 8*). By a method presently to be discussed, the logarithmically-amplified microphone signal is displayed on an oscilloscope in polar coordinates as a function of direction and at constant frequency.

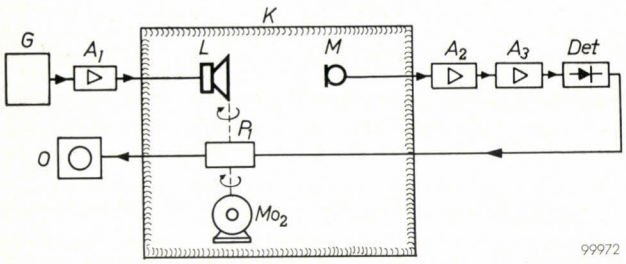


Fig. 8. Set-up for obtaining directivity patterns. *G* signal generator. A_1 push-pull amplifier. *L* loudspeaker under test in dead room *K* and, together with the liquid potentiometer P_1, rotated about a vertical axis by motor Mo_2. *M* microphone. A_2 microphone amplifier. A_3 logarithmic amplifier. *Det* detector. *O* oscilloscope.

We shall first explain how a polar diagram can be presented on an ordinary oscilloscope with two pairs of deflection plates. When a voltage $E \sin \omega t$ is applied to one pair of plates and a voltage $E \cos \omega t$ to the other pair, a circle is traced on the screen [5]). If these voltages are modulated in amplitude by a periodic time-function of fundamental frequency k. $\omega/2\pi$ (k an integer), a "modulated circle" is produced, that is to say one in which the radius is varied. The angle ωt is kept constantly equal to the angle α through which the loudspeaker is rotated during the time interval from 0 to t, and the amplitudes E are modulated in accordance with the logarithmically amplified microphone signal.

[5]) It is assumed for the sake of simplicity that the two pairs of plates are equally sensitive. The difference in sensitivity that actually exists must be corrected by adjusting the amplitudes of the two voltages to the correct ratio.

The voltages $E \sin \omega t$ and $E \cos \omega t$ are taken from a liquid potentiometer, the principle of which is illustrated in *fig. 9*. An insulating container *1* is filled with a conductive liquid (ethylene glycol). On opposite sides of the container carbon plates (*2*, *3*) are mounted, and between them a disc *4* of insulating material is rotated at a constant angular velocity ω. Fitted to the periphery of the disc, 90° apart, are two contacts of graphite (*5*, *6*) which are led out via slip rings. Connected between plates *2* and *3* is a centre-tapped battery of voltage $2E$, the centre-tap being earthed. The liquid in the plane of symmetry between the plates is therefore at earth potential, and the voltages on *5* and *6* vary, with respect to earth, as $E \sin \omega t$ and $E \cos \omega t$, respectively (disregarding a constant factor equal to the ratio of the diameter of the disc to the distance between the carbon plates).

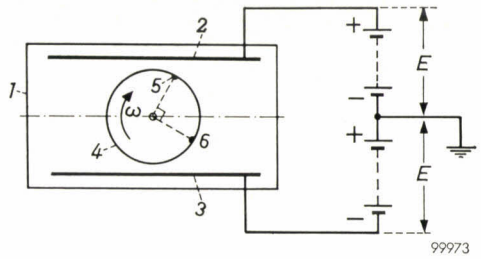
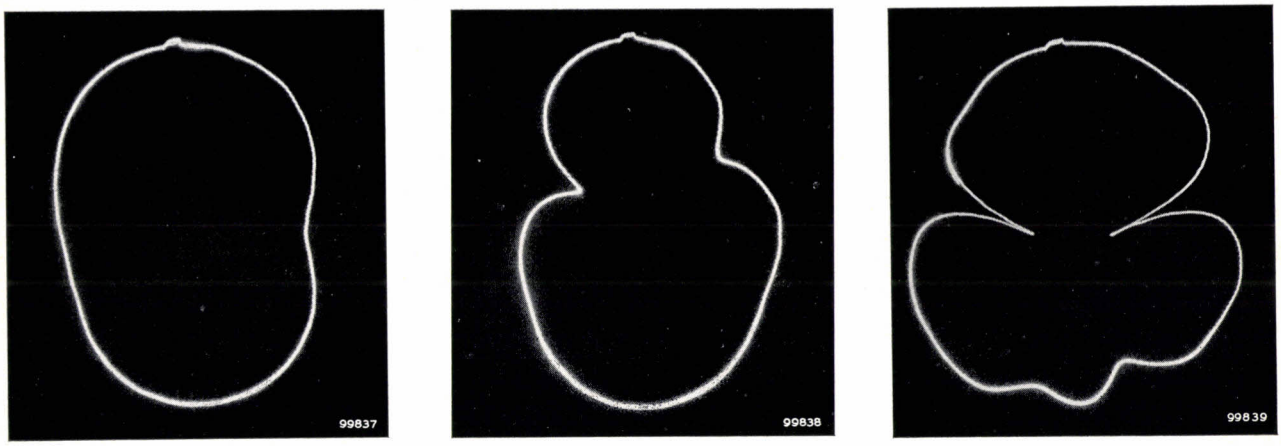
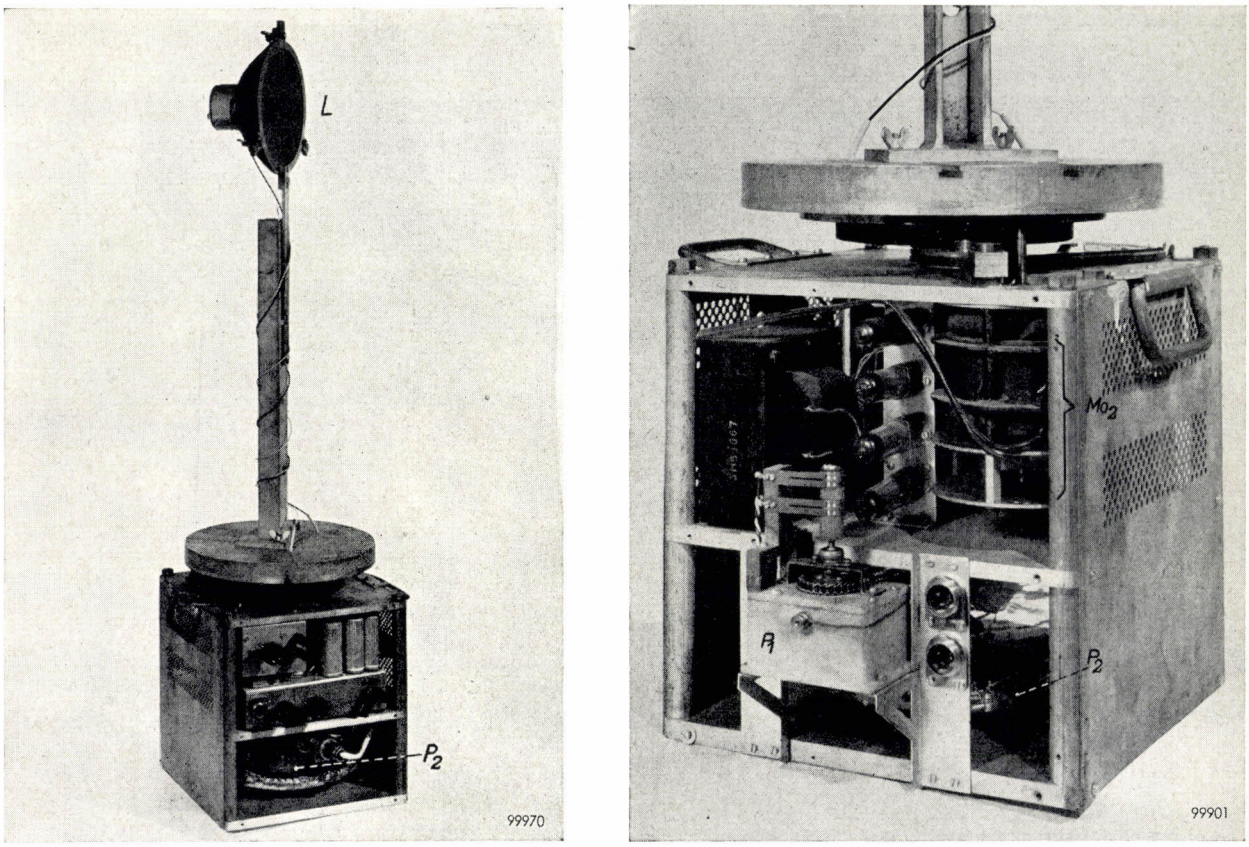


Fig. 9. Liquid potentiometer (P_1 in fig. 8) seen from above. *1* insulating container filled with a conducting fluid. *2*, *3* carbon plates. *4* disc of insulating material, rotating with angular velocity ω. *5*, *6* graphite contacts 90° apart on the periphery of the disc *4*. When voltages $+E$ and $-E$ with respect to earth are applied to plates *2* and *3*, the potentials of contacts *5* and *6* vary according to $E \sin \omega t$ and $E \cos \omega t$, respectively.

When these two voltages are applied to the deflection plates of the oscilloscope, a circle is traced on the screen which, when the applied voltage $2E$ is varied, becomes a "modulated circle". This varying DC voltage is obtained by rectifying the



1000 c/s 2000 c/s 3000 c/s



a *b*

Fig. 10. *a*) Mechanism for rotating the loudspeaker L about a vertical axis ($^1/_3$ rev per sec for recording directivity patterns). P_2 potentiometer required for recording efficiency curves. *b*) Mechanism seen from the other side, revealing the liquid potentiometer P_1 (cf. figs. 8 and 9) and the driving motor Mo_2.

output voltage of the logarithmic amplifier A_3 (fig. 8) by the detector *Det.*

The rotary mechanism, with a loudspeaker mounted on it, is shown in *fig. 10a.* The liquid potentiometer (P_1) can be seen in fig. 10*b*. *Fig. 11* reproduces some examples of directivity patterns recorded with the set-up described.

The logarithmic gain is obtained by using type EBF 80 pentodes in the first two stages of A_3. A feature of this valve is that the logarithm of the transconductance S at constant screen-grid voltage is approximately a linear function of the control-grid voltage (*fig. 12*). The rectified output signal is used for negatively biasing both EBF 80 pentodes. As a result, the logarithm of the gain is virtually proportional to the amplitude of the input signal, so that the microphone signal is recorded on

8000 c/s 16 000 c/s

Fig. 11. Directivity patterns recorded at different frequencies on a double-cone loudspeaker, type 9710 M (outer diameter 216 mm), without cabinet or baffle.

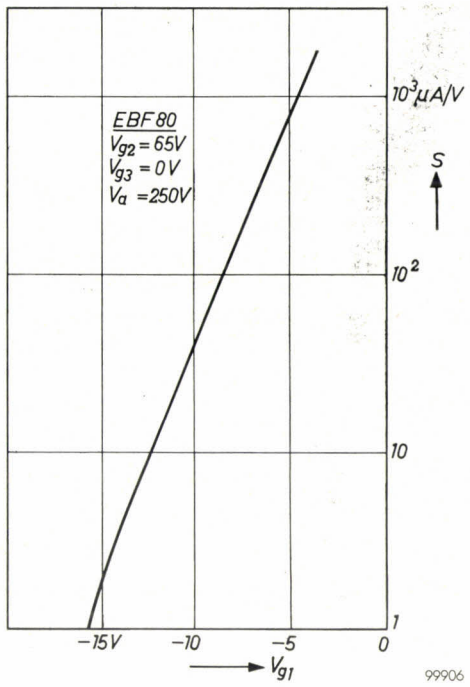


Fig. 12. Transconductance S (on logarithmic scale) of a pentode type EBF 80 as a function of control-grid voltage V_{g1}, at a screen-grid voltage of 65 V, a suppressor-grid potential of zero volts and an anode potential of 250 V. The characteristic is approximately straight, so that S is virtually a logarithmic function of V_{g1}.

a logarithmic scale.

The detector *Det* consists of two full-wave rectifiers, the output from one being positive and from the other negative with respect to earth. Full-wave rectifiers are used to reduce ripple and to provide effective smoothing. The ripple voltage must be kept small because it is superimposed as a spurious effect on the modulation of the circle. On the other hand, there is a limit imposed on the time constant for the smoothing: if it is too long, the direct voltages cannot vary rapidly enough and the pattern loses detail. The maximum permissible time constant is determined by the speed at which the loudspeaker is rotated (here $^1/_3$ revolution per second) and by the smallest angle of rotation within which the maximum pressure difference in the pattern must still be reasonably perceptible. This angle was taken as 5°, and the maximum permissible time constant is therefore: $(^5/_{360}) \times 3 = {}^1/_{24}$ sec. With full-wave rectification adequate smoothing is obtained with this fairly short time constant.

The oscilloscope

The transients and the directivity patterns are displayed on two oscilloscope tubes connected in parallel. One of these serves as the oscilloscope proper and must have a very long persistence screen, the time taken to trace a directivity pattern being 3 sec. The other tube is used for photographing oscillograms and is accordingly provided with a blue-luminescent phosphor of short persistence. Focusing and brilliance for the two tubes are controlled separately. A photograph of the oscilloscope rack appears in *fig. 13*.

For the recording of directivity patterns, both components of the signal voltage are modulated with the loudspeaker rotational frequency of $\frac{1}{3}$ c/s. AC amplifiers would have to be given impractically large *RC* values in order to reproduce such low-frequency signals without distortion. DC amplifiers are therefore used, the principal data for which are as follows:

Gain . . . . . 2000 ×,
± 1 dB from 0 to 20 kc/s,
± 3 dB to 50 kc/s.
Max. output voltage . 60 V.
Distortion . . . . . < 2%.
Input impedance . . approx. 0.5 MΩ.

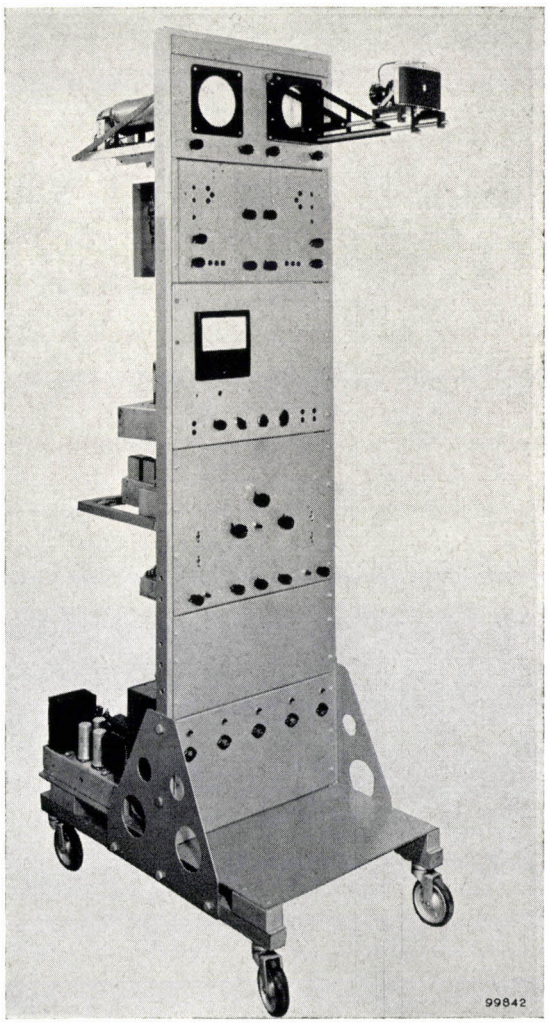


Fig. 13. Rack with two oscilloscope tubes and associated equipment. The left tube serves for observation of the directivity patterns and has a greenish-yellow luminescent phosphor of long persistence. The right tube is used when oscillograms are to be photographed, and has a blue-luminescent screen of short persistence.

The study of transients calls for a sawtooth sweep voltage of very low frequency, in order to be able to display at least one complete wave train plus the interval between two trains. The time involved may be up to 64 cycles of the audio frequency used. At the lowest audio frequency (20 c/s) the lowest sweep frequency must therefore be $^{20}/_{64} \approx 0.3$ c/s. The time-base generator uses a Miller integrator based on a transitron circuit [6]). This produces a closely linear sawtooth voltage of very low frequency without the need for a high supply voltage or high capacitances. The sawtooth voltage is amplified by one of the DC amplifiers mentioned above.

Loudspeaker efficiency as a function of frequency

The efficiency η of a loudspeaker is the ratio of the radiated acoustic power W_0 to the electrical input power W_i. The latter is the sum of W_0 and the losses, such as the electrical loss I^2R in the coil and the mechanical losses in the centering ring and in the rim of the cone. In the region of 500 c/s the efficiency of most loudspeakers is only a few per cent [7]); the losses there are thus much greater than W_0, and by far the largest part of these losses is due to I^2R. It is therefore the usual practice to measure efficiency at constant current I, and to regard W_i to the first approximation as identical with I^2R, although this is not correct in three respects:

1) it neglects the other losses with respect to I^2R,
2) it neglects W_0 with respect to I^2R,
3) it takes no account of the fact that R is frequency-dependent.

These errors only become significant, however, at frequencies higher than about 1000 c/s, mainly because the iron losses (or the copper losses if the loudspeaker is fitted with a short-circuiting ring) are then no longer negligible in relation to I^2R. At these frequencies, then, it is necessary to apply to the recorded curve a correction as a function of frequency. This correction takes account of the extra losses and also, in the region of several kc/s, of the radiated power. We shall not discuss here the determination of the correction curve, which can be done by the so-called three-voltmeter method.

Let us now turn to the method of determining the radiated acoustical power W_0. We assume that the loudspeaker possesses rotational symmetry about its axis. The sound intensity at a distance r from the loudspeaker, and in a direction α with respect to the axis, will be denoted by $I(\alpha)$. The power dW_0 radiated between the directions α and $\alpha + d\alpha$ (see *fig. 14*) is then:

$$dW_0 = I(\alpha) \times 2\pi r \sin\alpha \times r\, d\alpha.$$

For $I(\alpha)$ we can write $P^2(\alpha)/\varrho c$, where $P(\alpha)$ is the r.m.s. sound pressure at a distance r in a direction α, ϱ is the density of the air and c is the velocity of sound in air. This gives us:

$$dW_0 = \frac{2\pi r^2}{\varrho c} P^2(\alpha) \sin\alpha\, d\alpha.$$

The total radiated power W_0 is therefore:

$$W_0 = \frac{2\pi r^2}{\varrho c} \int_0^\pi P^2(\alpha) \sin\alpha\, d\alpha. \quad . \; . \quad (1)$$

If we have the directivity pattern of the loudspeaker at various frequencies, we could derive W_0 at these frequencies from the bounded area of the curve obtained by multiplying the radius vectors in

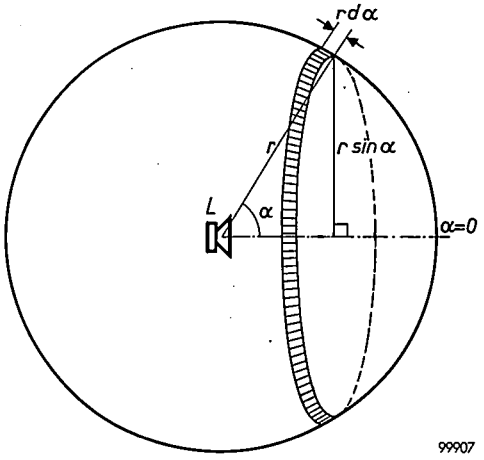


Fig. 14. When the loudspeaker L possesses rotational symmetry about the axis $\alpha = 0$, a sound intensity $I(\alpha)$ prevails at every point of the hatched segment. The surface area of the segment is $(2\pi r \sin\alpha) \times r\, d\alpha$. The acoustic power incident on the segment is $I(\alpha) \times 2\pi r^2 \sin\alpha\, d\alpha$. By integrating this expression over α (see eq. 1), we find the total radiated acoustic power W_0.

the directivity pattern by $\sqrt{|\sin\alpha|}$. This area can be found by planimetry, but this is a cumbersome method, particularly if it is required to find the efficiency for a whole series of frequencies.

In principal it would be possible to measure W_0 by disposing a very large number of microphones in a circle around the loudspeaker; by squaring the microphone voltages, multiplying them by the corresponding value of $|\sin\alpha|$ and adding the results together, we should find the approximate value of the integral in (1). However, this method is ruled out too by the high costs of the numerous condenser microphones that would be needed.

Fig. 15 illustrates a method by which, with only one microphone, the loudspeaker efficiency can be recorded directly as a function of frequency. The microphone is set up in a dead room, and in front

[6]) See e.g. F. Kerkhof and W. Werner, Television, Philips Technical Library, 1952, page 138 *et seq.*

[7]) J. de Boer, The efficiency of loudspeakers, Philips tech. Rev. 4, 301-307, 1939.

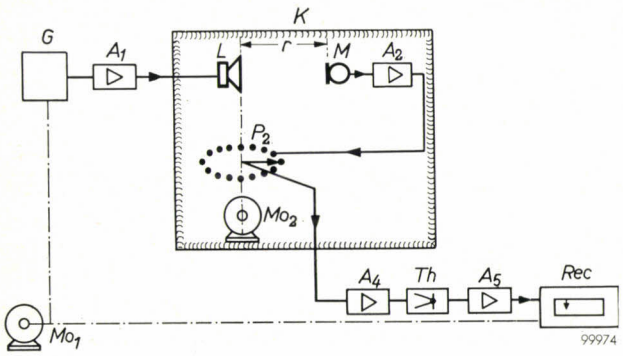


Fig. 15. Set-up for recording the efficiency of a loudspeaker as a function of frequency. L loudspeaker under test, fed from signal generator G via push-pull amplifier A_1. K dead room. M microphone. A_2 microphone amplifier. P_2 rotary potentiometer whose arm, like the loudspeaker, is rotated by the motor Mo_2 at about 3 revs/sec; the potentiometer has a $\sqrt{|\sin \alpha|}$ law. A_4 amplifier. *Th* thermocouple. A_5 DC amplifier. *Rec* recorder. The motor Mo_1 causes the frequency f of the signal generator to traverse the whole audio range in about 5 minutes, and at the same time drives the strip recorder at a corresponding rate.

of it the loudspeaker is rotated uniformly about a vertical axis. The angle α between the geometrical axis of the loudspeaker and the fixed connecting line between loudspeaker and microphone thus passes periodically through all values from 0 to 2π. To the output of the microphone amplifier a potentiometer P_2 is connected, whose arm rotates together with the loudspeaker. The potentiometer is so wound as to give a $\sqrt{|\sin \alpha|}$ law. The design of this potentiometer, which must be entirely free from crackle, is illustrated in *fig. 16*; see also fig. 10*a* and *b*. The output voltage of the potentiometer is applied to an amplifier which feeds the heater of a vacuum thermocouple. The direct voltage delivered by the thermocouple is amplified and recorded on a strip-chart recorder. The movement of the paper is coupled with the rotation of the capacitor in the signal generator which feeds the loudspeaker (just as in the case of the frequency response, see above). The amplifier between the signal generator and the loudspeaker operates in this case with negative current feedback, so that the current through the loudspeaker remains just about constant throughout the frequency range swept.

The fact that the deflection of the recording stylus is proportional to the radiated power, so that the efficiency is directly recorded as a function of frequency, can be explained roughly as follows. The proportionalities involved are:

Microphone voltage	$\propto P(\alpha)$,
Input voltage, potentiometer . .	$\propto P(\alpha)$,
Output voltage, potentiometer . .	$\propto P(\alpha)\sqrt{\|\sin \alpha\|}$,
Alternating current, thermocouple	$\propto P(\alpha)\sqrt{\|\sin \alpha\|}$,
Direct voltage, thermocouple . .	$\propto P^2(\alpha)\|\sin \alpha\|$.

$P^2(\alpha) \sin \alpha$ appears in (1) under the integral sign. The integration over α from 0 to π is effected by the rotation of the loudspeaker. The thermal inertia of the thermocouple ensures that the direct voltage obtained is proportional to the mean square value

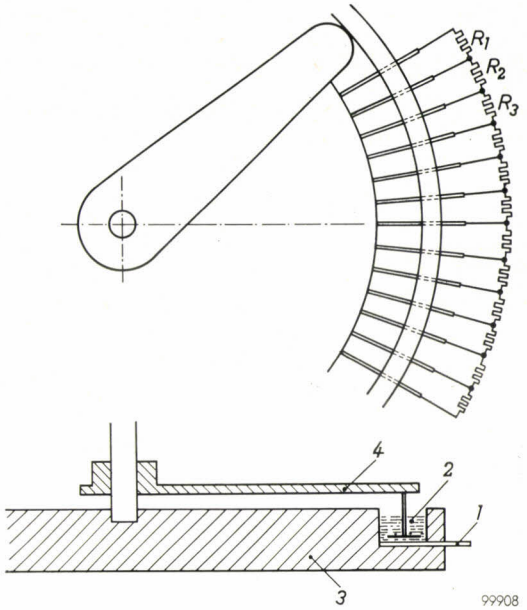


Fig. 16. Construction of the rotary potentiometer P_2 (fig. 15). $R_1, R_2, \ldots$ fixed resistors with values giving the potentiometer a $\sqrt{|\sin \alpha|}$ law. The taps are formed by 72 strip contacts *1* which are secured in the insulating plate *3* along the annular groove *2*. The groove is filled with a conducting liquid (ethylene glycol). The rotary arm *4* is clear of the strips but makes electrical contact with them via the conducting liquid. This construction guarantees complete freedom from crackle. The resistance values R are chosen low enough to make the conduction between the successive strips via the liquid negligibly low compared with the conduction through R. The fairly high contact resistance between the arm and the strips is of no consequence in that no current is drawn.

of the alternating current, which varies with α. A further condition is that the frequency should not change much during a single revolution of the loudspeaker. *Fig. 17* gives an example of an efficiency curve, which took about 5 minutes to record.

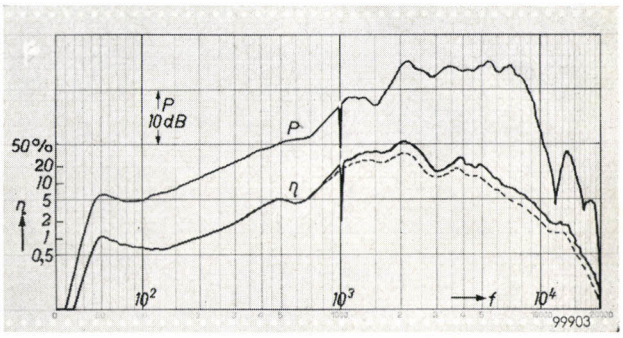


Fig. 17. Efficiency η (on logarithmic scale) recorded with the set-up in fig. 15; dashed curve: after correction for other losses than I^2R and for the radiated power. The recording was made in 5 minutes. For comparison, the frequency response (P) of the same loudspeaker is also shown.

When the loudspeaker is rotated through an angle α with respect to the position of the microphone, the instantaneous value $p(\alpha)$ of the sound pressure near the microphone is

$$p(\alpha) = P(\alpha)\sqrt{2}\sin(2\pi ft + \varphi),$$

where f is the sound frequency and φ is a phase angle.

Let the sensitivity of the microphone, including the microphone amplifier A_2, be s. The output voltage v_2 of A_2 is then:

$$v_2 = s\,P(\alpha)\sqrt{2}\sin(2\pi ft + \varphi).$$

The rotating potentiometer introduces a factor $\sqrt{|\sin\alpha|}$. Its output voltage, v_p, is therefore:

$$v_p = s\,P(\alpha)\sqrt{2\,|\sin\alpha|}\sin(2\pi ft + \varphi).$$

The thermocouple produces a direct voltage E_{th} which is proportional to the mean square value of v_p over half a revolution. Let the proportionality factor be ϑ, then:

$$E_{th} = \frac{\vartheta}{\pi}\int_0^{\pi} v_p^2\,d\alpha = \frac{2\vartheta s^2}{\pi}\int_0^{\pi} P^2(\alpha)\sin\alpha\sin^2(2\pi ft+\varphi)\,d\alpha. \quad (2)$$

Now we have

$$\sin^2(2\pi ft + \varphi) = \frac{1-\cos 2(2\pi ft+\varphi)}{2}.$$

Hence (2) becomes:

$$E_{th} = \frac{\vartheta s^2}{\pi}\int_0^{\pi} P^2(\alpha)\sin\alpha\,d\alpha - \frac{\vartheta s^2}{\pi}\int_0^{\pi} P^2(\alpha)\sin\alpha\cos 2(2\pi ft+\varphi)\,d\alpha.$$

It is easy to see that the second integral will be approximately zero, for even the lowest audio frequency f (20 c/s) is still high compared with the number of revolutions made by the loudspeaker per second, viz. 3. This means that during half a revolution ($\alpha = 0 \rightarrow \pi$) the factor $\cos 2(2\pi ft + \varphi)$ swings to and fro many times between $+1$ and -1, thereby making this integral virtually zero. This also applies if f is gradually raised during half a revolution. To a good approximation, then, only the first integral remains:

$$E_{th} \approx \frac{\vartheta s^2}{\pi}\int_0^{\pi} P^2(\alpha)\sin\alpha\,d\alpha.$$

In conjunction with (1) this gives:

$$W_o \approx \frac{2\pi r^2}{\varrho c}\frac{\pi}{\vartheta s^2}E_{th} = \frac{2\pi^2 r^2}{\varrho c\vartheta s^2}E_{th},$$

which demonstrates the approximate proportionality between W_o and E_{th}.

We have tacitly assumed above that the wave front is spherical. If it is not spherical, the contribution of the second integral in (2) will, in general, not be so small. It can be calculated, however, for a dipole — and as such we can regard a loudspeaker at low frequencies — which turns at two revolutions pers econd, that at $f = 80$ c/s the contribution of the second integral is smaller than 1%.

Other methods of measuring W_o

Two other methods of measuring W_o may be briefly mentioned. One resembles the measurement of the luminous flux of a light source in an integrating photometer: loudspeaker and microphone are placed in an acoustically "hard" room (with reflecting walls) in such a way that the direct radiation from the loudspeaker to the microphone is negligible. Since the wavelengths of the sound are relatively long, fairly complicated measures are necessary to prevent the formation of standing waves and to bring about a sufficiently diffuse distribution of the sound. This subject has already been dealt with in this review [8]).

The other method of measuring W_o was devised by the British Broadcasting Corporation [9]). As in our case, a potentiometer is made to rotate together with the loudspeaker, but the potentiometer law here is $|\sin\alpha|$. Use is further made of a somewhat modified kilowatt-hour meter. To the current coil of this meter a current is applied which is proportional to the microphone voltage, hence to $P(\alpha)$; to the voltage coil a voltage is applied which is taken from the potentiometer and is proportional to $P(\alpha)\,|\sin\alpha|$. In accordance with the principle of the kilowatt-hour meter the number of revs per sec of the disc in this meter is thus proportional to $P^2(\alpha)\,|\sin\alpha|$. The integration over α from 0 to π is again effected by rotating the loudspeaker. Since the kilowatt-hour meter only works reliably at frequencies of about 50 c/s, the microphone voltage must first be detected and the resultant direct voltage converted into an alternating voltage of 50 c/s. The method is not suitable for recording.

Calibration

It only remains to describe how the scale of the recorder is calibrated in milliwatts.

The microphone is disconnected from the amplifier A_2 (fig. 15) and the input of A_2 is connected to a signal generator with an output millivoltmeter. With the potentiometer P_2 rotating we can now determine the proportionality factor β between the square of the input voltage V_i of A_2 and the deflection x of the recorder:

$$x = \beta V_i^2. \quad (3)$$

The sensitivity s_m of the microphone is known, i.e. the proportionality factor between the sound pressure P and the voltage V_i:

$$V_i = s_m P. \quad (4)$$

All that we need now is the relation between the sound pressure P and the power W_o radiated by the loudspeaker. This relation can be calculated for a point sound source. Let the radiated power from this point source be W_o (uniformly distributed in all directions), then at a distance r from the source the sound intensity is $I = W_o/4\pi r^2$, and the relation between sound pressure P and I at that distance is given by $I = P^2/\varrho c$. Hence:

$$P^2 = \frac{\varrho c}{4\pi r^2}W_o. \quad (5)$$

---

[8]) R. Vermeulen, The testing of loudspeakers, Philips tech. Rev. 4, 354-363, 1939, in particular page 361.

[9]) A. Gee and D. E. L. Shorter, An automatic integrator for determining the mean spherical response of loudspeakers and microphones, B.B.C. Engng. Div. Monogr. No. 8, Aug. 1956.

From (3), (4) and (5) we find:

$$W_0 = \frac{4\pi r^2}{\varrho c\, s_m{}^2 \beta} x. \quad \ldots\ldots \quad (6)$$

In (6) the following are known:

r = distance from loudspeaker to microphone,
ϱ = density of the air (1.2 kg/m³ at 1 atm pressure and 22 °C),
c = velocity of sound (345 m/s at 22 °C),
s_m = sensitivity of the microphone, and
β = measured ratio between x and $V_i{}^2$.

With these known quantities we can therefore calculate the proportionality factor between W_0 and x.

In conclusion it may be noted that the equipment described can also be used *mutatis mutandis* for similar measurements on objects other than loudspeakers. Without any fundamental modifications it can serve, for example, for recording the frequency response and transients of amplifiers, and the directivity patterns of microphones, light sources and short-wave aerials.

Summary. Methods of testing loudspeakers are described, in which the sound pressure, the electrical impedance and the efficiency are recorded as a function of frequency (20-20 000 c/s) on a strip-chart recorder, whilst the directivity patterns at different frequencies and the transients produced by the sudden interruption of the electrical signal are displayed on a cathode-ray oscilloscope. The resonance frequency of the loudspeaker can also be determined. For most of the tests the loudspeaker and microphone (condenser type) are set up in an acoustically dead room. For recording the directivity patterns and the efficiency curve the loudspeaker is rotated about a vertical axis. About 5 minutes are required to record the efficiency curve.

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 Laboratories, Eindhoven, Netherlands.

2723: K. van Duuren and G. J. Sizoo: The gas-discharge mechanism for the argon-alcohol proportional counter (Appl. sci. Res. **B 7**, 379-399, 1959, No. 5).

Accurate measurements of the gas multiplication factor were done for a number of proportional counters with various wire diameters ranging from 0.003 to 0.1 cm and for various gas pressures ranging from 67 to 220 mm Hg. From these results the ionization coefficient for the argon-alcohol gas mixture 9 : 1 was deduced. It appeared that in a few cases the equilibrium between the mean electron energy and the electrical field was slightly disturbed and the conditions for the non-equilibrium situation could be established from the experiments. Rose's and Korff's calculations of the gas multiplication factor in proportional counters were compared with the authors' experimental results, whereby rather large discrepancies were found. Comparison of the authors' experimental values of the ionization coefficient for argon-alcohol with the earlier published values for pure argon and alcohol gives and insight in the mechanism of the electron multiplication process in the argon-alcohol gas mixture 9 : 1.

2724: J. M. Stevels and A. Kats: Défauts de réseau dans le quartz cristallin et la silice fondue (Compte-rendu XXXIe Congrès int. Chim. industr., Liège, Sept. 1958, Vol. II, pp. 125-128, Mercurius, Antwerp). (Lattice defects in quartz and fused silica; in French.)

In the Si-O lattice of quartz and of fused silica, defects — either foreign atoms or vacancies — are almost always present. The former are often not detectable by classical chemical methods owing to their low concentration. By study of dielectric losses, absorption spectra and paramagnetic resonance it is possible to gain information on the chemical nature of these imperfections.

2725: K. van Duuren and J. Hermsen: Improved design for halogen-quenched end-window Geiger counters (Rev. sci. Instr. **30**, 367-368, 1959, No. 5).

Note reporting certain improvements in the design of end-window Geiger-Muller tubes. The improvements are concerned with modifying the electric field especially near the window, and consist of the use of a conducting window and of a spherical or hemispherical anode.

2726: N. W. H. Addink and L. J. P. Frank: Remarks apropos of analysis of trace elements in human tissues (Cancer **12**, 544-551, 1959, No. 3).

It is shown that unless neoplasms develop in tissues relatively rich in zinc, the zinc level of both serum and whole blood in carcinoma cases is subnormal. There is no connection between the zinc level of whole blood on the one hand and the iron (in hemoglobin) or water level on the other. Control of 3 groups of carcinoma patients by measuring the zinc content of their whole blood during the years 1955 to 1957 resulted in agreement between the physician's view and that of the laboratory studying zinc level alterations, in 80 to 90% of the cases. The authors' conclusions are: (1) Blood from cancerous patients generally shows a subnormal zinc level (with a mean value of 4.5 parts per million as compared to a value of 6.5 parts per million for blood from healthy subjects). Favourable progress against the disease is accompanied by an increase towards normal values. In the case of unfavourable progress, the low zinc level is persistent or it decreases still further. (2) In cases of tumor development in tissues relatively rich in zinc (the aberrant tumor groups), the zinc level has been found to be supernormal. A fall to the normal zinc level indicates favourable progress, whereas a permanent supernormal level indicates unfavourable development in the illness. Shortly before death, supernormal as well as subnormal values can be found.

2727: J. Hornstra: Models of grain boundaries in the diamond lattice, I. Tilt about <110> (Physica **25**, 409-422, 1959, No. 6).

Small-angle grain boundaries are known to consist of arrays of dislocations. In this paper it is shown that in the diamond lattice not only for small, but for all angles of tilt about $[1\bar{1}0]$ a dislocation model can be constructed. There are no abrupt changes of model for small variations of the angle of tilt and the {111} twin boundary comes out automatically. This twin boundary can equally well be described as a twist boundary with a network of screw dislocations, but both models are identical with the generally accepted structure of the {111} twin boundary.

2728: M. J. Sparnaay: Van der Waals forces and fluctuation phenomena (Physica **25**, 444-454, 1959, No. 6).

An expression is given for the attraction between two electrically neutral systems each consisting of electric point charges which move at random inside spherical volumes. The calculation is based on the use of mean square values of density fluctuations in each system. The result and the method used are compared with Keesom's expression for the interaction between two dipoles and with the expression which is obtained for the interaction of two classical harmonic oscillators a large distance apart.

2729: P. H. J. A. Kleijnen: Travelling-wave-buizen (T. Ned. Radiogenootschap **24**, 71-88, 1959, No. 2/3). (Travelling-wave tubes; in Dutch.)

Fundamentals of the theory of the various forms of travelling-wave tube. The article substantially reproduces an introductory lecture given at the symposium on travelling-wave tubes organized by the Netherlands Radio Society in 1958.

2730: C. T. de Wit: Een 10-watt-lopendegolfbuis voor de 7,5-cm-band (T. Ned. Radiogenootschap **24**, 89-100, 1959, No. 2/3). (A 10-watt travelling-wave tube for the 7.5 cm waveband; in Dutch.)

Discussion of the main factors involved in the design and construction of a travelling-wave tube of high beam efficiency. An outline is given of the design of some of the components of the tube and of the circuit.

2731: A. Versnel: Lopendegolfbuizen met een laag ruisgetal (T. Ned. Radiogenootschap **24**, 101-112, 1959, No. 2/3). (Travelling-wave tubes with a low noise factor; in Dutch.)

The behaviour of current and velocity fluctuations in a travelling-wave tube are discussed. The calculations are done in 3 stages: first fluctuations between cathode and anode are considered, then the behaviour of these fluctuations in the drift space between anode and helical electrode and finally their behaviour when they interact with the signal injected on the helix. If the beam diameter, beam velocity in the drift space and beam current are kept constant, it is found that the noise factor depends only on the cathode-anode potential and on the length of the drift space. By a suitable choice of these parameters the minimum noise figure can be achieved. With certain assumptions, the calculated values agree with those measured on special laboratory tubes. It is shown that if certain conditions could be fulfilled it should be possible to achieve even lower noise factors.

2732: J. Ubbink: Masers, I (T. Ned. Radiogenootschap **24**, 129-136, 1959, No. 2/3; in Dutch).

Introductory article on the operation of the maser. Processes such as absorption and stimulated and spontaneous emission are discussed in connection with the inversion of energy-level population densities. This inversion (higher population in the higher

levels), often characterized by the term "negative temperature", is essential to maser operation. The reason why magnetic dipole systems are more favourable than electric dipole systems is outlined. It is also explained why low temperatures are favourable to maser operation. The product of amplification and bandwidth and the noise of a maser are discussed with the aid of an equivalent circuit.

2733: T. Kralt, H. D. Moed, E. J. Ariëns and Th. W. J. Hendriksen: Synthesis and pharmacological properties of O-acetylsalicylamides (Rec. Trav. chim. Pays-Bas **78**, 199-206, 1959, No. 3).

The synthesis of a series of N-substituted O-acetylsalicylamides is described. Most compounds possess weak antipyretic activity and strong spasmolytic activity.

2734: T. Kralt, H. D. Moed, E. J. Ariëns and Th. W. J. Hendriksen: Synthesis and pharmacological properties of salicylamides (Rec. Trav. chim. Pays-Bas **78**, 207-214, 1959, No. 3).

The synthesis of a series of N-substituted salicylamides is described. Most compounds possess weak antipyretic activity and strong spasmolytic activity.

2735: A. Venema: The production of ultra-high vacua by means of a diffusion pump (Vacuum **9**, 54-57, 1959, No. 1).

The result of a measurement of the so-called ultimate pressure of a diffusion pump is often not only determined by the pump itself, but also by the additional apparatus necessary for making this measurement. An analysis is given of the gases which may be present at the pump mouth and the factors determining the pressure of those gases are treated. A discussion of the methods which can be used for making these pressures as small as possible results in a design of a pump apparatus. A description of this apparatus, which includes a mercury diffusion pump with cooling traps, is given. With the apparatus a pump speed at the recipient of 5-10 l/s for the common gases is obtained and an ultimate pressure lower than 10^{-12} mm Hg is reached. (See also Philips tech. Rev. **20**, 145-157, 1958/59.)

R 387: J. van Laar and J. J. Scheer: Photoelectric determination of the electron work function of indium (Philips Res. Repts. **15**, 1-6, 1960, No. 1).

The photoelectric work function of evaporated indium films prepared in ultra-high vacuum is found to be 4.08 ± 0.01 eV.

R 388: C. Kooy and U. Enz: Experimental and theoretical study of the domain configuration in thin layers of $BaFe_{12}O_{19}$ (Philips Res. Repts. **15**, 7-29, 1960, No. 1).

With the aid of the optical Faraday effect the domain configuration and the magnetization process in thin transparent single-crystal plates of $BaFe_{12}O_{19}$ are studied. The plates have surfaces parallel to the basal plane of the hexagonal structure. The domain pattern consists of line-shaped domains. The domain width is measured as dependent on the applied magnetic field parallel to the c-axis. The width of a reversed domain decreases slowly with the applied field reaching a finite thickness near saturation, whereas the width of a domain magnetized in the direction of the applied field increases rapidly near saturation. In this stage the remaining reversed line-shaped domains contract towards cylindrical domains which collapse at slightly higher fields. Saturation is reached in a field well below $4\pi I_s$. The demagnetizing energy of a partly magnetized thin uniaxial crystal having the easy axis normal to the surface is calculated for a domain pattern consisting of straight parallel domains. The stable domain configuration for a given value of the magnetic field is obtained by minimizing the total energy. The solution is obtained in the form of two simultaneous equations containing the two different domain widths, which are evaluated by an electronic computer. Theoretical magnetization curves are deduced. The general accordance between experiment and theory is good.

R 389: K. Teer: Investigations into redundancy and possible bandwidth compression in television transmission (Philips Res. Repts. **15**, 30-96, 1960, No. 1).

Continuation of **R 385**.

R 390: M. E. Wise: On the radii of five packed spheres in mutual contact (Philips Res. Repts. **15**, 101-106, 1960, No. 2).

These results form part of a study of close packing in spheres. Each of four given spheres touches the other three. The fifth ("interstitial") sphere touches all four externally and is usually surrounded by them. The other sphere touching all four either lies outside them or encloses them. These properties are shown by geometrical inversion. In this way a general equation relating to the radii of five spheres in mutual contact is derived. Some general conclusions about different configurations of the five spheres are discussed on the basis of this relation. The formulae are also expressed in terms of statistical functions

of the four radii; the volume of the tetrahedron of centres of the first four spheres is shown to be related to the same functions.

R 391: F. van der Maesen: Determination of numbers of injected holes and electrons in semiconductors (Philips Res. Repts. **15**, 107-119, 1960, No. 2).

In semiconductors, deviations Δn and Δp of the equilibrium numbers n_0 and p_0 of electrons and holes are unequal in many cases because of trapping. It is shown how measurements of the photo Hall-effect and photoconduction may give information on the numbers Δn and Δp separately. From the ratio Y of the relative change of the Hall-effect and of the photoconductivity, the quantity $K = \overline{\Delta n/\Delta p}$ can be evaluated. Graphs of Y versus K for some substances are given. It is shown how K occurs in the formulae used in computations of the diffusion-recombination length L from photoelectromagnetic and photoconductive data. If r_n and r_p denote the ratios of the Hall-mobility to drift mobility for electrons and holes, respectively, r_n/r_p can be evaluated from the measurements where $K = 1$. In this way the value of r_n/r_p at room temperature is found to be 0.54 ± 0.05 for germanium and 0.78 ± 0.04 for solicon.

R 392: C. A. A. J. Greebe and W. F. Knippenberg: Grown *p-n* junctions in silicon carbide (Philips Res. Repts **15**, 120-123, 1960, No. 2).

An account is given of the preparation and properties of grown junctions in SiC. The forward characteristics are tentatively explained on the basis of a *p-i-n* structure. *P-n* luminescence has been observed containing violet light.

R 393: M. T. Vlaardingerbroek: Small-signal performance and noise properties of microwave triodes (Philips Res. Repts. **15**, 124-221, 1960, No. 2).

This paper (thesis, Eindhoven 1959) reports on a theoretical and experimental study of the behaviour of a microwave triode at 4 Gc/s. After a general survey of the field, the equivalent circuit of the triode is deduced in section 2. The internal feedback is accounted for by a series-resonant circuit in the grid lead. The noise behaviour is described by a noise-current source and a noise-voltage source connected to the input terminals of the equivalent circuit. In section 3, those elements of the equivalent circuit which result from the presence of electrons in the active space and the noise sources are calculated on the basis of the single-velocity transit-time theory of Llewellyn and Petersen. Four phenomenological constants are introduced, to account for the effects of the electrons returning in front of the potential minimum and for the multi-velocity effects on the original "Rack" velocity fluctuation and the convection-current fluctuation at the cathode surface. The methods of measurement necessary for the determination of the fourpole coefficients of the triode and its four characteristic noise quantities, giving the magnitudes of the noise sources and their cross-correlation, are described in section 4. These methods are complicated because the distance between the measuring devices and the object under study is of the order of several wavelengths. In section 5 the experiments are described and the results are compared with the theory of sections 2 and 3. The conclusions are: (1) The electronic admittances of the triode can be calculated with reasonable accuracy on the basis of the single-velocity transit-time theory provided that the current density is so high that the distance from cathode to potential minimum is small compared to the cathode-grid distance. (2) The noise properties of the triode are only slightly affected by total-emission noise, transit-time spread and feedback. The contribution of reflection of electrons at the anode surface to the noise properties of a microwave triode is of some importance and can be estimated. (3) The random emission from the cathode is the most important noise source. The four measured noise quantities can be used for calculating the four phenomenological noise quantities introduced in section 3. From the values obtained for the latter quantities, the minimum noise figure and the beam-noise parameters S and Π of an electron-beam amplifier equipped with an identical cathode can be calculated. The results are in reasonable agreement with the measured values of S and Π.

R 394: G. Meijer and M. Avinor: Excitation spectra of vanadium-activated zinc and cadmium sulphide and selenide phosphors (Philips Res. Repts. **15**, 225-237, 1960, No. 3).

Excitation spectra were measured for the 2 μ fluorescence band of vanadium-activated zinc and cadmium sulphide and selenide phosphors. The emission is excited by absorption in two composite bands due to vanadium at 1.1 and 1.6 eV, by absorption in an auxiliary impurity centre, such as copper or silver, if present, and by fundamental excitation.

R 395: W. van Gool and A. P. Cleiren: Self-activated and Cu-activated fluorescence of ZnS (Philips Res. Repts. **15**, 238-253, 1960, No. 3).

Two series of experiments on the fluorescence of ZnS are described. The first one, of which only a review of the experimental results is given, presents some additional data on the theory of ZnS activated with Cu, made some years ago by Kröger and coworkers. In the second series a limited number of phosphors have been studied in order to settle a few questions. In particular, the difference between the blue Cu emission and the blue self-activated emission of ZnS has been studied. Furthermore, the temperature dependence of the fluorescence bands and the influence of the coactivator on both mentioned fluorescences and the green copper fluorescence were determined. The results can be interpreted by assuming that the low-temperature fluorescence is dependent on the coactivator. The room-temperature emission bands can have a composite character, in such a way that in addition to the low-temperature emission band another band may be important. The results are to some extent uncertain, due to experimental difficulties. These have been discussed and it is stressed that further careful experimental work may be more important for our knowledge of the ZnS fluorescence than detailed calculations about some special model.

R 396: W. van Gool, A. P. Cleiren and H. J. M. Heijligers: Fluorescence of some activated ZnS phosphors (Philips Res. Repts. **15**, 254-274, 1960, No. 3).

Several series of ZnS phosphors were prepared in H_2S atmosphere at 1150-1200 °C. Activators used were Ag, Cu, Au and coactivators were Al, Sc, Ga, In. Phosphors were made with all combinations of activators and coactivators with one concentration. In other phosphors equal concentrations of selected pairs of activators and coactivators were studied at different levels. Some special series of phosphors were made in addition, and spectral distributions of all phosphors at room temperature and at —196 °C are reported. The spectral distributions can be separated into parts of low and high photon energy. The high-energy parts can be attributed to the presence of the activators. The low-energy parts are of a composite structure. Part of these bands is due to an associate centre of activator and coactivator. Other parts of the low-energy bands may be due to other centres, which could not be identified unambiguously.

R 397: J. A. W. van der Does de Bye: Measurement of decay times of excess carriers in semiconductors, excited by X-ray pulses (Philips Res. Repts. **15**, 275-289, 1960, No. 3).

A method for the measurement of transient decay times of excess carrier concentrations in homogeneous semiconductors is described. The exciting agent is X-rays, which are delivered in short pulses of 0.1-0.3 μsec. A reasonable excitation (about 10^{12} cm^{-3} per pulse in Ge) requires peak currents of several tens of amperes through the X-ray tube, which is fitted with a large dispenser cathode. Two pulse generators, which can deliver short, high-tension pulses of 80 kV and of 150 kV, are used. These two voltages imply two different effective linear X-ray attenuation coefficients for every substance. For germanium they are about 20 cm^{-1} and 8 cm^{-1}. This affords a reasonable possibility of exciting the bulk without much disturbance by surface recombination. Excitation, dissipation, sample geometry and noise set a minimum to the resistivity of the sample for the production of a clearly visible decay curve on the oscilloscope. For Ge this is between 0.1 and 1 ohm cm. The measuring apparatus comprises an exponential time base and a decay-curve simulator for the measurements of decay time constants, down to 0.1 μsec. Measurements performed on copper-doped germanium yield decay times which are also found with chopped light. Some measurements on CdTe are also described.

R 398: H. G. Grimmeiss and H. Koelmans: *P-n* luminescence and photovoltaic effects in GaP (Philips Res. Repts. **15**, 290-304, 1960, No. 3).

GaP crystals were prepared from the elements. Crystals made at low phosphorus pressure mainly showed *n*-conductivity with an activation energy of 0.07 eV. Crystals with *p*-conductivity were obtained by heating at high phosphorus pressure (activation energy 0.19 eV) or by doping with Zn. The non-doped crystals showed electroluminescence in bands at 6250 Å and 5650 Å. The electroluminescence is shown to be due to the recombination of charge carriers, within *p-n* junctions via levels within the forbidden gap. A level scheme for undoped GaP is proposed. The crystals showed point-contact rectification and photovoltaic effects. On measuring the photovoltage as a function of wavelength, excitation bands were found at 4200 Å and 5600 Å in non-doped crystals and at 4200 Å and 6000 Å in Zn-doped crystals. The long-wave excitation peaks of the photovoltage are explained with a two-step mechanism, one optical and one thermal.